ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

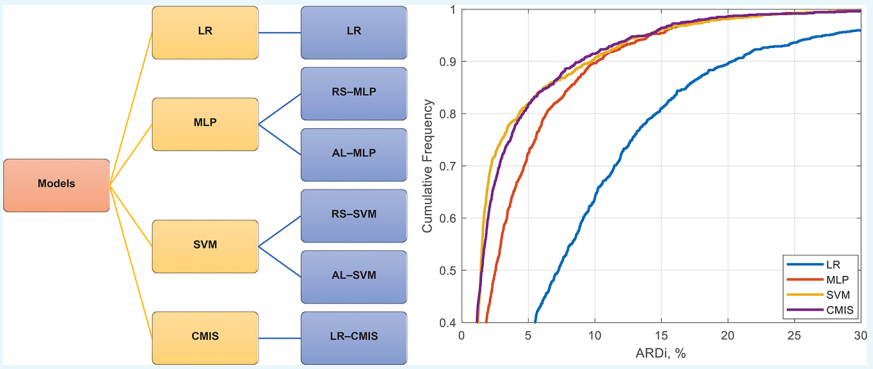
Data-Driven Connectionist Models for Performance Prediction of Low Salinity Waterflooding in Sandstone Reservoirs

Afshin Tatar, Ingkar Askarova, Ali Shafiei,* and Mahsheed Rayhani

Cite This: *ACS Omega* 2021, 6, 32304−32326

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Low salinity waterflooding (LSWF) and its variants also known as smart water or ion tuned water injection have emerged as promising enhanced oil recovery (EOR) methods. LSWF is a complex process controlled by several mechanisms and parameters involving oil, brine, and rock composition. The major mechanisms and processes controlling LSWF are still being debated in the literature. Thus, the establishment of an approach that relates these parameters to the final recovery factor (RF_f) is vital. The main objective of this research work was to use a number of artificial intelligence models to develop robust predictive models based on experimental data and main parameters controlling the LSWF determined through sensitivity analysis and feature selection. The parameters include properties of oil, rock, injected brine, and connate water. Different operational parameters were considered to increase the model accuracy as well. After collecting the relevant data from 99 experimental studies reported in the literature, the database underwent a comprehensive and rigorous data preprocessing stage, which included removal of duplicates and low-variance features, missing value imputation, collinearity assessment, data characteristic assessment, outlier removal, feature selection, data splitting (80−20 rule was applied), and data scaling. Then, a number of methods such as linear regression (LR), multilayer perceptron (MLP), support vector machine (SVM), and committee machine intelligent system (CMIS) were used to link 1316 data samples assembled in this research work. Based on the obtained results, the CMIS model was proven to produce superior results compared to its counterparts such that the root mean squared rrror (RMSE) values for both training and testing data are 4.622 and 7.757, respectively. Based on the feature importance results, the presence of Ca^{2+} in the connate water, Na^{+} in the injected brine, core porosity, and total acid number of the crude oil are detected as the parameters with the highest impact on the RF_f. The CMIS model proposed here can be applied with a high degree of confidence to predict the performance of LSWF in sandstone reservoirs. The database assembled for the purpose of this research work is so far the largest and most comprehensive of its kind, and it can be used to further delineate mechanisms behind LSWF and optimization of this EOR process in sandstone reservoirs.

1. INTRODUCTION

Extraction of hydrocarbon from mature reservoirs is a challenging topic since only 30%−50% of the OOIP can be recovered through primary to secondary stages.[1] Thus, producing part of the oil left in the reservoirs is of paramount importance. Several EOR strategies from sandstone reservoirs have been proposed and implemented in both experimental and field trials.[2−7] Among these approaches, waterflooding and subsequently low salinity waterflooding (LSWF) remain an attractive option.[8,9] This method first gained recognition in the 1990s by research works conducted by Morrow and co-workers.[8,10−13] Several researchers have attempted to delineate mechanisms behind the low salinity effect (LSE) by exploring the impact of each parameter on the results.[14−19] The mechanisms of the LSE, which are categorized as brine/rock, oil/brine, and oil/brine/rock groups,[20] can be listed as follows:

Received: October 3, 2021
Accepted: November 3, 2021
Published: November 16, 2021

wettability alteration,[21] multicomponent ion exchange,[18] fines migration,[13] local pH increase,[17] double layer expansions (DLE),[22,23] osmotic effect,[24] microdispersion formation,[25] and viscoelastic effect.[26] By reviewing the literature, one can conclude that each of these mechanisms is directly or indirectly dictated by several effective parameters relevant to the oil, brine, and rock composition.[27] Hence, exploring the relationship between these parameters and the mechanisms of the LSE and, in turn, the final recovery factor (RF_f) is a superior task for researchers in this area.

To this aim, several experimental studies[28−31] as well as modeling studies[23,32−34] have been carried out. Austad et al.[17] proposed that multiple ion exchange between Ca^{2+} on clay sites with H^+ from the injected brine induces a local pH increase, which is a cause of wettability alteration and higher oil recovery. In addition, acid/base reactions between crude oil and OH^- of the invaded brine also assist in desorption of polar oil components from the rock surface. The increase in the pH gradient due to the adsorption/desorption processes is supported by the next research studies conducted by the same group.[35,36] Such observations have led to the establishment of some conditions as the backbones of the LSE. These conditions include the presence of clay particles, polar oil components, and potential-determining ions (i.e., Ca^{2+}, Mg^{2+}, and SO_4^{2-}) in the connate water.[17,37] However, some of these conditions, such as the necessity of the presence of clay, have been questioned since many works have observed the LSE in the absence of clay.[38−42] In support of the DLE mechanism, Nasralla and Nasr-El-Din[43] conducted several ζ-potential experiments with various brine salinities and ion types (Na^+, Mg^{2+}, and Ca^{2+}). The results indicated the significant role of Na^+ in altering the surface charge at the oil/brine and brine/rock interfaces, which was similar to higher oil recovery results with sodium-rich brine compared to Mg^{2+} and Ca^{2+}. In line with that, Yang et al.[44] also observed higher oil recovery with NaCl (1509 ppm) compared to $CaCl_2$ (961 ppm) brine with similar ionic strength. They pinpointed that the presence of Ca^{2+} generates strong bonds with polar oil components on the rock surface, which induces retention of polar oil components by lowering oil detachment. As per oil composition, the literature provides strong bodies of evidence regarding the impact of total acid number (TAN) and total base number (TBN) of crude oils on oil production results.[45,46] In this regard, McMillan et al.[47] proved the critical role of oil composition in the wettability of the rock surface. They observed that crude oils with heavy fractions (i.e., asphaltene and resin) are in favor of observing the LSE rather than crude oils with lower heavy ends. Results regarding lower oil recovery with crude oils of higher TBN are reported by RezaeiDoust et al.[48] and Fjelde et al.[49] They concluded that the presence of basic components causes direct crude oil drop pining on the clay surface, increasing the retention, and hence impairing wettability alteration.

Similarly, many other experimental works aimed at shedding light on mechanisms and conditions controlling the performance of the LSWF.[50−54] Nonetheless, conventional experimental investigations imply that studies under laboratory conditions suffer from some issues including time and budget constraints, uncertainty in the quality of the data obtained and their reliability, the requirement of highly skilled personnel to design and conduct laboratory experiments, and some other common limitations of experimental research such as slight differences in standards and procedures. In recent years, owing to accumulation of a considerable amount of experimental data including coreflooding experiments in sandstone reservoirs, application of machine learning (ML) and artificial intelligence (AI) has attracted interest in petroleum engineering research ranging from EOR to flow assurance.[55−63] As an example, AI-based models are used to estimate distribution of unwanted shale barriers in the SAGD procedure[62] or provoke the acquisition of real-time water saturation from readily available well log data.[63] Hassan et al.[64] provided simple yet effective models utilizing artificial neural network (ANN), support vector machine (SVM), and fuzzy logic system (FLS) approaches to determine the acid fracturing impact exerted on the productivity of naturally fractured carbonates. It showed that the FLS is a powerful predictive tool for well stimulation, reservoir geomechanics, and permeability evaluation aspects. Furthermore, AI models, especially the ANN, assisted in foreseeing the capacity of propped hydraulic fractures to conduit fluids in gas shale zones.[65,66] In line with that, Artun[67] utilized the ANN to recognize the connectivity between injector and producer wells using neural network weights into the features. In addition, the capability of ML methods in reservoir simulation was assessed by Ghassemzadeh et al.[68] through incorporating an extensive range of reservoir data and creating a deep net simulator (DNS). The developed model was able to address some issues in this area such as improving slow computational time, which is common in conventional physics-based simulators. Al-Gawfi et al.[69] used genetic programming to develop three generalized correlations for the prediction of thermophysical properties of solvent/bitumen mixtures such as viscosity, density, and solubility with implications in the design of thermal- and solvent-based bitumen recovery processes. The scientists used a comprehensive set of experimental phase behavior data to develop and validate the correlations.

Despite the extensive application of AI in petroleum and chemical enginering, very few research works have investigated the performance of the LSWF in petroleum reservoirs via developing intelligent predictive models.[20,70−72] Different predictive tools, such as the ANN, adaptive neural fuzzy inference system (ANFIS), SVM, decision tree (DT), and random forests (RF), assist in the estimation of LSWF potential. Apart from these methods, other approaches, like the least squares support vector machine (LSSVM), which is a modified version of the SVM and extra tree (ET), are also proposed and utilized.[71] The developed ML solutions are required to be accurate, reliable, and vigorous to allow objective assessment of the LSWF performance. Several statistical parameters including mean squared error, coefficient of determination, and relative deviation aid in verification of the developed models.[72] Based on the data-driven analysis of LSWF[20] reported by Wang et al.,[70] several models were built such as the multiple linear regression (MLR), ANN, SVM, and RF to conduct a comparative analysis and present an optimum model with the best performance. The tertiary recovery factor, which ranges from 0 to 16.2%, was chosen as the output variable that has to be predicted. In total, 178 points were collected for model development, from which 80% trained and 20% tested the models. The models were then optimized to obtain reliable results. Thus, hyperplanes were optimized in the SVM to get the maximum distance between data points. The number of trees in the RF has ranged from 1 to 100 trees. Also, in the ANN model, the number of hidden layers changed from one to two with the number of neurons changing from 1 to 12.

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

Input variables were divided into five subsets, each of which covered certain interactions and were selected based on the MLR results performed previously.[20] The RF approach demonstrated the highest R^2 with an average value of 0.728 and was selected by the researchers as the optimum model. The SVM approach showed the highest correlation between brine and rock conditions and the recovery factor. Kondori et al.[71] applied a data-driven approach to assess the LSWF performance. Three models, the LSSVM, ANFIS, and ET, were built. Six input variables were selected to forecast the recovery factor during low salinity waterflooding under secondary oil recovery mode. They chose the rate of low salinity brine injection, temperature, oil viscosity, rock permeability, porosity, and brine salinity as parameters that affect oil recovery. Overall, 257 data points were utilized, 90% of which assisted in training models and the remaining were left for testing. The reliability of models was evaluated by relative error, mean squared error, mean average percentage error, and R^2. The trial-and-error approach was then applied to identify the optimum depth and number of the trees. In the described study, a maximum depth of 24 and 11 trees were used. The R^2 values reached 0.995 at the training stage and 0.934 at the testing stage. Among the applied data-driven approaches, the highest mean squared error was observed in the ANFIS model. This trend was explained by the complexity of the model in choosing the correct membership function types. The LSSVM assisted by the coupled simulated annealing (LSSVM−CSA) approach resulted in good performance with a total R^2 of 0.988. The salinity of the injection brine was found to have the highest impact on the recovery factor with the importance of 29.12%. Recently, Hidayat and Astsauri[72] used RF regression to determine the significant parameters of LSWF and to establish a correlation between the parameters and the RF_f. Using 1000 design of experiment parameters by CMOST from CMG software, the temperature of reservoir and injected brine, volume of injected brine, formation water, and injection water composition from carbonate reservoirs were considered. Based on the obtained results from the RF algorithm, the concentration of SO_4^{2-} in both formation water and injection water followed by injection volume appeared to be the most significant parameters in designing LSWF. However, these are simulation results and not verified yet by any experimental data. Mean squared error was used for sensitivity analysis. The R^2 score for five random state variations was reported to be more than 0.9.

A close look into the literature implies that previously developed smart models were developed based on a relatively limited number of data despite the abundance of a notable amount of experimental data in this subject area. In addition, to reach a reliable predictive AI/ML-based intelligent model, larger databases are required to overcome some deficiencies presented in the current models. Owing to the great number of experimental works published in credible journals on the performance of LSWF in sandstone reservoirs on the core scale now, it is possible to assemble a much larger database. Despite the fact that the amount of uncertainty associated with the data grows with growing the size of database due to some shortcoming and inconsistencies in conducting some of the experiments, a larger database allow the ML approaches to learn more about the process. To this end, 1316 experimental data points from 99 experimental studies reported in credible journals and conference proceedings on the performance of LSWF in sandstone reservoir rocks were collected. By considering the RF_f as the target parameter, four ML algorithms, namely, the linear regression (LR), multilayer perceptron (MLP), SVM, and committee machine intelligent system (CMIS), were used for building new smart models. To increase the model accuracy and reliability, removing low-variance features, data imputation, assessment of data collinearity, removal of outliers and duplicates, feature importance, splitting, and scaling data were applied as preprocessing stages. In addition, to assist with the process of selecting the best model, several graphical and statistical approaches were also used.

This paper is organized in four main parts. In the second section (Methodology), Model Development, Data Gathering, and Data Preprocessing, specifically Removing Duplicates and Low-Variance Features, Missing Value Imputation, Collinearity Assessment, Data Characteristics, Outlier Removal, Data Scaling, and Hyperparameter Optimization, are described in detail. In the third section, the performance of the model with assessment of several statistical and graphical error functions accompanied with the leverage method are discussed. Finally, in the Results and Discussion section, feature selection, selected model's accuracy, outlier diagnosis, and EOR implications are discussed.

2. METHODOLOGY

In this section, data preprocessing stages, hyperparameter optimization techniques, different modeling techniques, and optimization algorithms used in this research including linear regression, multilayer perceptron neural network, support vector machine, committee machine intelligent system, and applicability domain concepts are described briefly.

2.1. Model Development. In this section, the development of the AI models is presented systematically. At first, root mean squared error (RMSE) (eq 11) was selected as the key parameter for the cost function. The 10-fold cross-validated RMSE is considered to check the performance of the developed models. Different predictive models used in this research work are shown in Figure 1. There are four families of predictive models optimized by different methods. The details are provided in the consequent sections.

For the hyperparameter optimization, a random seed number is fixed for the training process. In other words, a particular set of hyperparameters results in a unique error value.

2.2. Data Gathering. Any reliable AI model can be developed only by using precise, true, and comprehensive data. Several parameters are introduced in the literature as parameters affecting the LSWF efficiency. The majority of these works are experimental studies. A thorough literature survey was conducted with the aim of not only including the available experimental data but also identifying the effective parameters in a coherent, logical, and technically sound methodology. In this regard, 99 studies[3,5,8,11−14,16,26,30,31,35,38,43−45,47,48,73−152] were inspected and analyzed carefully for the purpose of this research work. Five general categories of rock (13 features), oil (eight features), brine (22 features), and connate water properties (22 features) as well as the operational parameters (five features) were recognized. Rock properties include the core physical properties and composition of the sandstone rocks, including length (l, cm), diameter (d, cm), porosity (Φ, %), brine permeability (K_b, mD), irreducible water saturation (S_{wi}, %), clay content (C_{Clay}, %wt), quartz content (C_Q, %wt), mica

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

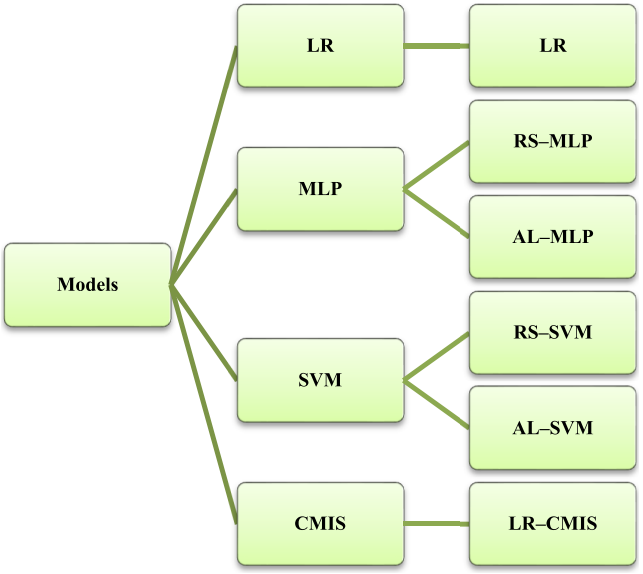


Figure 1. Different methods developed and tested in this research work. The MLP and SVM will be subjected to the hyperparameter optimizations of AL and RS.

content (C_{Mic}, %wt), calcite content (C_{Cal}, %wt), dolomite content (C_{Dol}, %wt), carbonate content (sum of calcite and dolomite contents) (C_{Car}, %wt), anhydrite content (C_A, %wt), and feldspar content (C_F, %wt). Density (ρ, g/cm^3), viscosity (μ, cP), total acid number (TAN, mg KOH/g), total base number (TBN, mg KOH/g), saturates content (Saturate, %), aromatic content (Aromatic, %), resin content (Resin, %), and asphaltene content (Asphaltene, %) represent the oil properties. Recorded in different temperatures, ρ and μ were not used directly; rather, the ratio of the mentioned parameters was used to the absolute temperature. This method will not impose any threat to our results since ρ and μ and the absolute temperature have a ratio level of measurement. The ionic concentrations were considered for both brine and connate water, which include bicarbonate (HCO_3^-, ppm), chloride (Cl^-, ppm), metaborate (BO_2^-, ppm), fluoride (F^-, ppm), nitrate (NO_3^-, ppm), bromine (Br^-, ppm), azide (N_3^-, ppm), sodium (Na^+, ppm), potassium (K^+, ppm), lithium (Li^+, ppm), sulfate (SO_4^{2-}, ppm), magnesium (Mg^{2+}, ppm), calcium (Ca^{2+}, ppm), barium (Ba^{2+}, ppm), strontium (Sr^{2+}, ppm), carbonate (CO_3^{2-}, ppm), ferrous (Fe^{2+}, ppm), dithionite ($S_2O_4^{2-}$, ppm), ferric (Fe^{3+}, ppm), total monovalent ions (MI, ppm), total divalent ions (DI, ppm), and salinity (S, ppm). The operational parameters are the controllable testing conditions set by the operator. The temperature (T, °C), injection flow rate (Q, mL/min), aging time (t, h), initial recovery factor, (RF_i, % OOIP), and the flooding stage (secondary or tertiary) were considered for the operational parameters. All of the considered parameters in this research work are numeral values except for the flooding stage, which is separated into secondary and tertiary stages. Either the final oil recovery (RF_f, % OOIP) or the incremental oil recovery (which is the difference between the RF_f and RF_i) can be used as the target parameters. We have selected the final oil recovery as the target parameter here. Despite the importance of RF_i, recent modeling papers have disregarded this parameter in their analysis. For example,

Table 1. Independent Variables Used in This Study[a]

	Operational Parameters	Rock Properties	Oil Properties	Brine Properties	Connate Water Properties
1	T, °C	l, cm	ρ, g/cm^3	HCO_3^-, ppm	HCO_3^-, ppm
2	Q, mL/min	d, cm	~~μ, cP~~	Cl^-, ppm	Cl^-, ppm
3	t, h	φ, %	TAN, mg KOH/g	~~BO_2^-, ppm~~	~~BO_2^-, ppm~~
4	RF_i, % OOIP	K_b, mD	TBN, mg KOH/g	~~F^-, ppm~~	~~F^-, ppm~~
5	Flooding Stage	S_{wi}, %	$Saturate$, %	~~NO_3^-, ppm~~	~~NO_3^-, ppm~~
6		C_{Clay}, %wt	$Aromatic$, %	~~Br^-, ppm~~	Br^-, ppm
7		C_Q, %wt	$Resin$, %	~~N_3^-, ppm~~	N_3^-, ppm
8		~~C_{Mic}, %wt~~	$Asphaltene$, %	Na^+, ppm	~~Na^+, ppm~~
9		~~C_{Cal}, %wt~~		~~K^+, ppm~~	K^+, ppm
10		C_{Dol}, %wt		~~Li^+, ppm~~	~~Li^+, ppm~~
11		~~C_{Car}, %wt~~		SO_4^{2-}, ppm	~~SO_4^{2-}, ppm~~
12		C_A, %wt		~~Mg^{2+}, ppm~~	Mg^{2+}, ppm
13		C_F, %wt		Ca^{2+}, ppm	Ca^{2+}, ppm
14				~~Ba^{2+}, ppm~~	Ba^{2+}, ppm
15				~~Sr^{2+}, ppm~~	~~Sr^{2+}, ppm~~
16				CO_3^{2-}, ppm	CO_3^{2-}, ppm
17				Fe^{2+}, ppm	Fe^{2+}, ppm
18				~~$S_2O_4^{2-}$, ppm~~	~~$S_2O_4^{2-}$, ppm~~
19				~~Fe^{3+}, ppm~~	~~Fe^{3+}, ppm~~
20				MI, ppm	MI, ppm
21				DI, ppm	~~DI, ppm~~
22				S, ppm	S, ppm

[a]The removed low-variance features are shown in this table.

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

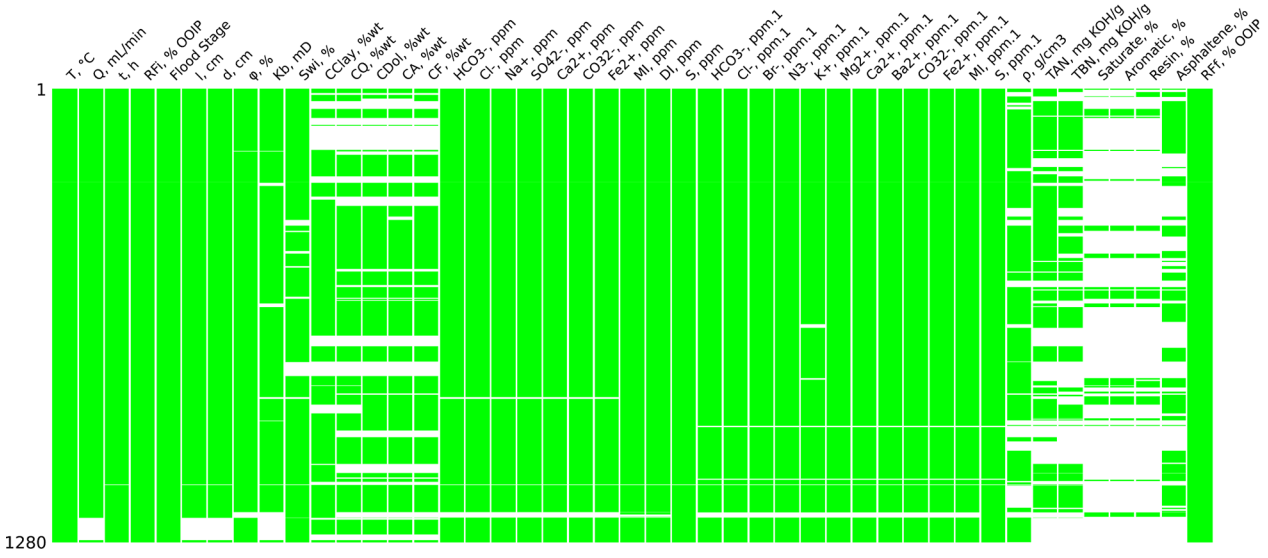


Figure 2. Depiction of missing values in the database. The empty sections of the database are shown as blank areas.

Kondori et al.[71] modeled recovery factors spreading from 33 to 93%. As a rule of thumb, this range implies the RF_f values as the target value. Neglecting RF_i could carry deceptive information such that the two cases with the RF_f equal to 50% but with various incremental oil recoveries can be interpreted alike. The same goes for the studies that only used incremental oil recoveries. For example, Wang and Fu[20] and Wang et al.[70] have used the incremental oil recoveries and provided no details about the RF_i.

2.3. Data Preprocessing. In this section, a comprehensive and rigorous approach is presented for data preprocessing. Processes such as data gathering, removal of duplicates and low-variance features, missing value imputation, collinearity assessment, data characteristics, outlier removal, feature selection for modeling, data splitting, and data scaling are described in detail.

2.3.1. Removing Duplicates and Low-Variance Features. Two main types of duplicates, namely, Type I and Type II, should be properly treated. Type I duplicates represent data samples for which both independent and dependent parameters are identical. For such cases, the procedure is to maintain the first instance and remove the rest. Type II duplicates, which are less common compared to its counterpart, identical independent variables, result in different target values. For these cases, although a set of variables can be selected as the proper one based on the domain experts' discretion, they are all removed in this research work. Low-variance features are parameters with a fixed value for most of the data samples. For example, for a particular phenomenon, all the tests (i.e., 1000 samples) are performed under room conditions but one study has reported the results for the same phenomenon at elevated temperatures (i.e., two samples). A feature having just one value is called a zero-variance feature. Such features do not make a positive contribution to the modeling. Hence, it is suggested to remove them to decrease the computation load. For this, a threshold value should be defined. It is worthwhile mentioning that the data should be normalized before calculating the variance as it is a distance-based property. Herein, the data are normalized in the range of [0 1] before calculating the variance. The low-variance features were determined as those having variances below 0.005 (99.5% of values are the same). Removing columns and rows might result in emerging duplicates or low-variance columns, respectively. Consequently, elimination should be done in a loop until no such undesired data remains.

In this process, three Type I and 33 Type II duplicates were identified and removed from our database. Table 1 shows the removed low-variance features from the database including no operational parameter, three rock properties, one oil property, 12 injected water properties, and 10 connate water properties. The crossed line denotes that the parameter is removed. Subsequently, the size of the database became 1280 rows × 45 columns, in addition to the target.

2.3.2. Missing Value Imputation. Usually, the experimental works lack some of the information reported in Table 1, such as the oil properties or the rock composition. The simplest method to remedy the missing values is to easily drop either the corresponding row or column, though this method will result in losing valuable information. This will negatively influence the ability of AI methods in learning as much as possible about the EOR process. Hence, missing data imputation can serve as the best option to prevent missing valuable information from the experimental data published in the literature. Imputation can be done using a middle point (mean, mode, or median) or by a modeling algorithm that predicts the missing values based on the other parameters. If the number of missing values is too high, imputation is not effective. Thus, the number of each feature should be checked before applying any imputation method. Missingno and Bilogur[153] used a suitable tool to depict the existing missing values. Figure 2 depicts the missing values in the database gathered for the current research work. As one can see, lots of missing values exist for some parameters, especially for oil properties. Imputation will be associated with high levels of uncertainty if the percentage of missing values exceeds a threshold value. Thus, dropping is the best treatment for such features. In this research work, a threshold of 50% was considered, based on which three features, Saturate, Aromatic, and Resin, with respective missing value percentages of 86.25, 86.25, and 85.47%, were removed from the database. Asphaltene with a missing percentage of 47.89% was deemed to survive. It should be mentioned here that the mean value

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

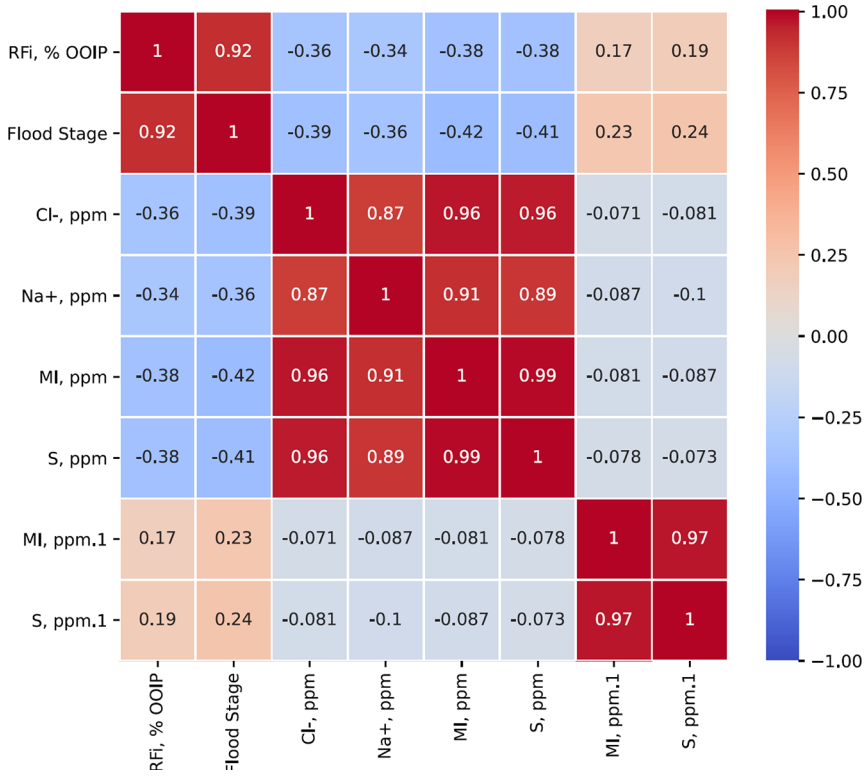


Figure 3. Collinear features in the database. The mentioned values depict the Pearson *R*. The color bar provides a graphical comparison among the *R* values.

Table 2. Independent Variables Used in This Study[a]

	Operational Parameters	Rock Properties	Oil Properties	Brine Properties	Connate Water Properties
1	T, °C	l, cm	ρ, g/cm^3	HCO_3^-, ppm	HCO_3^-, ppm
2	Q, mL/min	d, cm	~~μ, cP~~	Cl^-, ppm	Cl^-, ppm
3	t, h	φ, %	TAN, mg KOH/g	~~BO_2^-, ppm~~	~~BO_2^-, ppm~~
4	RF_i, % OOIP	K_b, mD	TBN, mg KOH/g	~~F^-, ppm~~	~~F^-, ppm~~
5	~~Flood Stage~~	S_{wi}, %	~~$Saturate$, %~~	~~NO_3^-, ppm~~	~~NO_3^-, ppm~~
6		C_{Clay}, %wt	~~$Aromatic$, %~~	~~Br^-, ppm~~	Br^-, ppm
7		C_Q, %wt	~~$Resin$, %~~	~~N_3^-, ppm~~	N_3^-, ppm
8		~~C_{Mic}, %wt~~	$Asphaltene$, %	Na^+, ppm	~~Na^+, ppm~~
9		~~C_{Cal}, %wt~~		~~K^+, ppm~~	K^+, ppm
10		C_{Dol}, %wt		~~Li^+, ppm~~	~~Li^+, ppm~~
11		~~C_{Carc}, %wt~~		SO_4^{2-}, ppm	~~SO_4^{2-}, ppm~~
12		~~C_A, %wt~~		~~Mg^{2+}, ppm~~	Mg^{2+}, ppm
13		C_F, %wt		Ca^{2+}, ppm	Ca^{2+}, ppm
14				~~Ba^{2+}, ppm~~	Ba^{2+}, ppm
15				~~Sr^{2+}, ppm~~	~~Sr^{2+}, ppm~~
16				CO_3^{2-}, ppm	CO_3^{2-}, ppm
17				Fe^{2+}, ppm	Fe^{2+}, ppm
18				~~$S_2O_4^{2-}$, ppm~~	~~$S_2O_4^{2-}$, ppm~~
19				~~Fe^{3+}, ppm~~	~~Fe^{3+}, ppm~~
20				~~MI, ppm~~	~~MI, ppm~~
21				DI, ppm	~~DI, ppm~~
22				~~S, ppm~~	S, ppm

[a]The removed collinear features are shown in this table.

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

Table 3. Statistical Parameters for Independent and Dependent Variables Used in This Research Work

parameter	minimum	Q_1	median	Q_3	maximum	mean	mode	s	k	range	IQR	SD
T, °C	15.55	25	60	75	148	58.77	60	0.3	−0.31	132.45	50	27.21
Q, mL/min	0	0.2	0.5	1	54.65	1.74	0.5	7.03	62.75	54.65	0.8	5.11
t, h	0	120	336	504	1680	348.25	336	0.96	1.39	1680	384	277.15
RF_i, % OOIP	0	0	28.95	53.49	88.2	28.04	0	0.27	−1.56	88.2	53.49	28.85
l, cm	1	7.1	7.67	12.83	50.8	11.61	7.6	2.71	7.82	49.8	5.73	9.43
d, cm	0.4	3.74	3.8	3.81	11.56	3.74	3.8	3.96	28.45	11.16	0.07	1.05
Φ, %	5.1	19	21	24	68	21.85	20	2.19	12.82	62.9	5	6.7
K_b, mD	0.1	50	121.93	466.47	16,205	466.47	466.47	9.13	102.61	16204.9	416.47	1292.12
S_{wi}, %	0	20.3	25.94	32	81.1	25.94	25.94	0.22	1.58	81.1	11.7	10.31
C_{Clay}, %wt	0	3	5	5.8	40	5.75	5.75	2.6	9.9	40	2.8	5.12
C_Q, %wt	0	84.67	86	90	100	84.67	84.67	−2.57	8.69	100	5.33	11.38
C_{Dob}, %wt	0	0	0	0.82	16	0.82	0	5.32	30.89	16	0.82	2.21
C_F, %wt	0	1	3	4.69	40	4.69	4.69	2.92	8.91	40	3.69	6.96
HCO_3^-, ppm	0	0	0	21.2	2060	124.98	0	3.47	13.42	2060	21.2	331.03
Cl^-, ppm	0	705.51	2767	14629.43	33333.5	7594.79	0	0.9	−0.49	33333.5	13923.92	8357.22
Na^+, ppm	0	496.74	1825	10,650	25,000	5178.28	0	1.05	−0.05	25,000	10153.26	6006.85
SO_4^{2-}, ppm	0	0	0	39.5	4664	327.25	0	3.06	8.84	4664	39.5	888.49
Ca^{2+}, ppm	0	0	53	425	16666.67	392.06	0	8.87	100.67	16666.67	425	1232.01
CO_3^{2-}, ppm	0	0	0	0	121.3	1.3	0	9.76	95.31	121.3	0	11.69
Fe^{2+}, ppm	0	0	0	0	1	0.03	0	5.78	31.53	1	0	0.16
DI, ppm	0	0	188.23	1414.17	16666.67	1148.44	0	3.22	15.79	16666.67	1414.17	2066.69
HCO_3^-, ppm	0	0	0	176	210,953	1898.6	0	11.18	123.31	210,953	176	18557.3
Cl^-, ppm	0	16,215	20504.78	35,976	233,400	35,976	22,214	2.29	6.4	233,400	19,761	37955.05
Br^-, ppm	0	0	0	0	1	0.05	0	4.17	15.4	1	0	0.22
N_3^-, ppm	0	0	0	0	50	1.11	0	6.54	40.87	50	0	7.31
K^+, ppm	0	0	55	350	14,474	514.49	0	4.18	20.95	14,474	350	1377.69
Mg^{2+}, ppm	0	24	281.6	1382	11,910	838.04	0	4.11	22.72	11,910	1358	1422.81
Ca^{2+}, ppm	0	159	545	3329.4	36,800	2920.58	0	3.51	14.08	36,800	3170.4	5910.33
Ba^{2+}, ppm	0	0	0	0	961.31	14.6	0	9.36	95.43	961.31	0	83.59
CO_3^{2-}, ppm	0	0	0	0	8610	74.21	0	11.06	121.45	8610	0	760.53
Fe^{2+}, ppm	0	0	0	0	1	0.01	0	8.56	71.68	1	0	0.11
S, ppm	118.6	30,000	39,000	92189.36	1,000,000	68379.07	40,000	4.87	43.08	999881.4	62189.36	80827.32
ρ, g/cm^3	2.02×10^{-3}	2.45×10^{-3}	2.68×10^{-3}	2.91×10^{-3}	3.30×10^{-3}	2.68×10^{-3}	2.68×10^{-3}	-3.29×10^{-1}	-6.06×10^{-1}	1.28×10^{-3}	4.59×10^{-4}	2.86×10^{-4}
TAN, mg KOH/g	0	0.18	0.85	0.92	9.07	0.85	0.85	5.59	41.9	9.07	0.74	1
TBN, mg KOH/g	0	1.16	1.45	1.45	3.87	1.45	1.45	0.38	0.42	3.87	0.3	0.58
Asphaltene, %	0	3.2	3.9	4	10.8	3.9	3.9	0.71	1.13	10.8	0.8	2.07
RF_p % OOIP	9.77	47	55.43	66.75	95.24	56.21	51	−0.24	−0.01	85.47	19.75	14.57

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

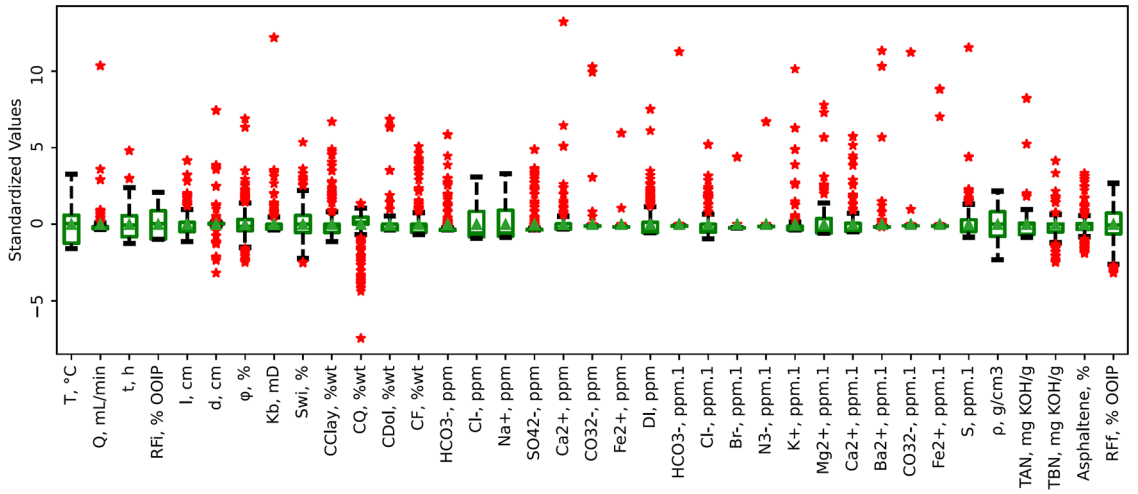


Figure 4. Box plot for the used data in this research work for the total database. The rectangular box and the red asterisks show the IQR and fliers, respectively.

was used for imputation in this research work. The database was checked for the low-variance features after imputation again. As a result, C_A was removed. After this stage, the database size became 1280 rows × 41 columns.

2.3.3. Collinearity Assessment. The presence of collinear independent parameters in a database is common, especially for those having a large number of independent parameters. Such features just increase the complexity of the model and do not make a positive contribution. The Spearman correlation factor (R) is the preferred option to assess the collinearity among the independent parameters for the non-normal distribution of the features' data. This parameter is assessed mutually between just two features. The correlation matrix heatmap is the best tool to depict the R values. A threshold value is defined in a way that values above the threshold indicate collinearity. In this research work, the threshold value is set at 0.9. In this sense, more than two parameters may be collinear and when they are recognized, just one should be retained, and the others should be dropped. The variance inflation factor (VIF) shows the multicollinearity among the independent values and can be used as the criterion for this purpose. Lower VIF values denote a better condition regarding multicollinearity. For the collinear features, the one with the least VIF value is retained, while the rest are dismissed. Figure 3 depicts the correlation matrix heatmap for the collinear features. As is shown, only the injected water features are involved with collinearity. For the connate water, two features are collinear, thus one should be selected. The VIF values for MI and S are 534.70 and 310.14, respectively. Based on the mentioned procedure, we kept the S and removed the MI. By removing the collinear features, the VIF values tremendously decrease. For example, this parameter for S underwent an about 60-fold decrease and reached 5.28. After this modification, the average for the VIF values decreased from 74.98 to 21.76. Table 2 shows the remaining and removed features during this process. Such reductions decreased the number of rows and columns in the database to 1280 rows × 37 columns.

2.3.4. Data Characteristics. Table 3 lists the statistical parameters of the used features. To have an easy and fast understanding of the data distribution, the box plot for each parameter is provided too. It is worthwhile mentioning that the standardized values are applied for this plot to secure our judgment upon scale diversity. After standardizing, the data vector represented standard deviation and average values of 1 and 0, respectively:

$$X^i_{\text{Standard}} = \frac{X^i_{\text{Original}} - \bar{X}_{\text{Original}}}{\text{SD}_{\text{Original}}} \quad (1)$$

Figure 4 depicts the box plot. In this plot, the green rectangle, green horizontal line, and red asterisks denote average, median, and fliers, respectively. The height of the blue box represents the interquartile range (IQR).

2.3.5. Outlier Removal. The outliers in the preprocessing stage are those data samples that are far from the majority of data, though they are not wrong. These data are called interesting outliers and should be studied in a separate devoted study. For example, Figure 4 suggests that there is at least one extreme value for C_Q, which is far different from the other data points. This point corresponds with two 0 values data samples reported by Al-Saedi et al.[145] Excluding these values, the lowest value for quartz content is 34.9%, which is a great gap in data. The inclusion of this point worsens the data distribution, diverting it from normality. Additionally, due to the lack of data for C_Q in a range of 0−34.9, the developed model will have unstable and poor results if tested for a data sample in this range. Hence, such points should be removed before modeling.

In this research work, a threshold limit equal to 9 times the standard deviation (SD) from the mean value was used in this regard. Through applying this method, 70 interesting outliers were removed from the database that correspond to 5.47% of the database, which is a reasonable amount. The database was checked for the low-variance features after outlier removal. At this step, Cl^- for the connate water was removed. Afterward, 1210 rows and 35 columns plus the target remained in the database.

2.3.6. Data Scaling. Data scaling accelerates optimization speed. Additionally, it prevents the undue effects of large-scale features. The standard scaling, eq 1, is used for the scaling of points. As a result, data will have a mean value equal to 0 and a standard deviation equal to 1. The scaler is first fitted on the training dataset and then is applied on the testing dataset. Using the scaled data, all the features, regardless of their different scale, would have the same effect on the modeling.

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

2.3.7. Hyperparameter Optimization. Each model contains some hyperparameters and parameters. The former is set before the training process and substantially affects model performance. Application of optimum values of hyperparameters guarantees maximum performance of the developed models. The basic way to find the best hyperparameters is through a trial and error process. This is done by changing the values and monitoring corresponding changes in the results. Nonetheless, this method is time-consuming and it fails to define the optimum values in most cases. Grid search is another method that discretely explores the searching space by examining all the possible combinations of predefined values. Another approach is the random search (RS), which does not have the limitation of discrete exploring. Hyperopt[154] is a Python library developed for hyperparameter optimization, which is suitable for both categorical and numerical searching spaces. Two optimization methods provided by this library, RS and annealing (AL), are used in this research work.

2.4. Data Visualization Using William's Plot. The graphical illustration of the applicability domain (AD) can be presented by William's plot, where the horizontal axis is for leverages and the vertical axis is for standardized deviation (SD_i):[155]

$$SD_i = \frac{D_i}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} D_i^2 \times (1 - h_i)}} \quad (2)$$

where D_i represents the deviation.

The cutoff value for leverages is h^* and for SD_i are ±3. Thus, if the data sample has $h_i < h^*$ and is bounded in the range of $-3 < SD_i < 3$, then it lies in the AD of the developed model and is referred to as good high-leverage. The data beyond that area are identified as outliers or bad high-leverage data.

3. COMPUTATIONAL PROCEDURE

An overall 1316 data samples including 70 independent variables are gathered from 99 research studies published in the literature. Several processing techniques were used for this database through which both the size and dimension of the database were decreased in an attempt for optimization. The standard training−testing technique is used in this research work. The training dataset is used to determine the model parameters. A 10-fold cross-validation will control the model for preventing overfitting. In other words, it shows the simulation's stability. The testing dataset is not introduced to the model during the training phase. It has no role in optimizing either model's hyperparameters or parameters. In this research work, the database is split into two subsets of training and testing in the ratio of 80:20. During the training process, the results obtained from the validation dataset determine which model passes to the next iteration. On the other hand, to compare the developed models, the accuracy for the testing dataset determines which model is the best in terms of accuracy. In addition, the applied models in this work include the linear regression (LR), multilayer perceptron neural network, support vector machine, and committee machine intelligent system. A brief description of these models is provided in the next section.

3.1. LR and Machine Learning Algorithms. The models used in this research work are provided by the Scikit-learn[156] library for python. A brief discussion of which is as follows.

3.1.1. Linear Regression. The LR is a simple model aiming at providing a predictive model by a simple linear equation. In this method, a constant is assigned for each parameter and a bias term is introduced. Hence, if there are m independent parameters, which are also called regressors, $m + 1$ coefficients should be determined. These coefficients will be determined using the ordinary least squared method. Clearly, this method is a white-box approach and is easy to interpret. An ordinary least squared method was used to determine the coefficients of the following equation:

$$\sum_{i=1}^{N} a_i \times P_i + a_0 \quad (3)$$

The scaled parameters are not used for this method. The 10-fold cross-validated RMSE was calculated and checked for 100 trials. The optimized coefficients and biases are listed in Table 4 by 10 significant figures. The best 10-fold cross-validated error was determined as 13.1773.

Table 4. Optimized Coefficients for the Developed LR Model

P_i	a_i
RF_i, % OOIP	0.174403184
l, cm	−0.103002000
TAN, mg KOH/g	−3.037204816
K_b, mD	−0.000759568
S_{wi}, %	−0.18244322
Φ, %	0.159497675
Q, mL/min	0.867920573
Na^+, ppm	0.000139149
Ca^{2+}, ppm	0.000249526
ρ, g/cm^3	363.239017
d, cm	2.997792449
a_0	41.75870739

3.1.2. Multilayer Perceptron. The objective of an ANN model is to investigate complex nonlinear relationships. Several ANNs differ from each other in various aspects. The most popular one is MLP, which is a layered-structure model with input, hidden, and output sublayers. Each of these sublayers has a certain number of neurons or nodes, where the nodes of one layer are interconnected with nodes from the other one. The connection to each node carries a weight, the value of which can be uniform, random, or predetermined.[157] The number of neurons in the input/output layers is in agreement with the number of input/output variables. The MLP networks are trained by optimizing the weights and biases for each connection via several iterations until the output values are close enough to the experimental ones. This process continues until the minimum errors and best predictive performance are obtained.

The Scikit-learn[156] library provides several hyperparameters for MLP. If the user does not set values, the default ones will be used for the model development. Regarding the size of the database for this research work, the most important hyperparameter is the number of hidden layers as well as the number of neurons in them. The Hyperopt[154] receives a range as the searching space. Both numbers of hidden layers and the contained neurons are set by one parameter. For example, (12) and (8, 6, 2) denote networks with one hidden layer containing 12 neurons and 3 hidden layers with 8, 6, and 2 neurons, respectively. A maximum of two hidden layers with overall neuron numbers of 30 was considered for this research

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

work. The solver was introduced manually as "lbfgs". A comparison between two different optimization algorithms for the hyperparameters is presented in Figure 5 and Table 5. In

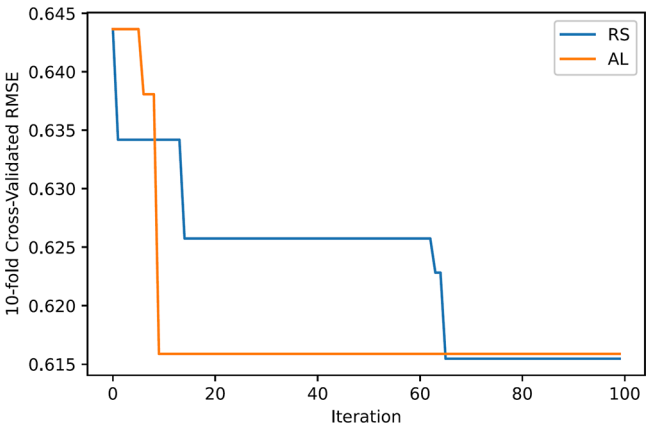


Figure 5. Convergence of the hyperparameters to the optimum state for the MLP model. RS has shown a better performance.

Figure 5, the vertical axis denotes the 10-fold cross-validated RMSE for the scaled data. As it is evident, the RS method has yielded superior results to the AL method.

Table 5. Optimized Values Acquired by Different Methods for the Developed MLP Model

	range	RS	AL
hidden layer size	(1,) to (30, 30)	(13, 18)	(17, 5)
best error		0.6155	0.6159

After determination of the optimum structure for the model, the fixed random seed number for the training was removed to find the best predictive model. This change enables different paths for the training. For this purpose, training was repeated for 100 iterations. The best error value was determined to be 0.5915.

3.1.3. Support Vector Machine. The support vector machine (SVM), which is developed based on the Vapnik–Chervonenkis theory, is a classifier that separates the data into two or several classes. The SVM deals with the generalization problem, which refers to the performance of the model with unseen data. In addition, this method works well with a dataset in which the number of features is more than the number of samples.[158,159]

The most two important hyperparameters for SVM are the regularization parameter (C) and kernel coefficient (*gamma*). The search space was introduced to the Hyperopt[154] to find the optimum values. For this problem, both the search spaces are of continuous type. The results of optimization are shown in Figure 6 and Table 6. As is shown, the AL has yielded a better error equal to 0.5769.

3.1.4. Committee Machine Intelligent System. Several techniques can be used in the AI modeling process to identify the most suitable model. In the end, only the model with the best performance is selected, while the rest are discarded. In this regard, the CMIS method has been developed to address this deficiency.[160] This model is grounded on the "divide and conquer" principle, which implies the splitting of complex tasks into several parts, which are solved separately. The results of each part are then combined and represent the solution for

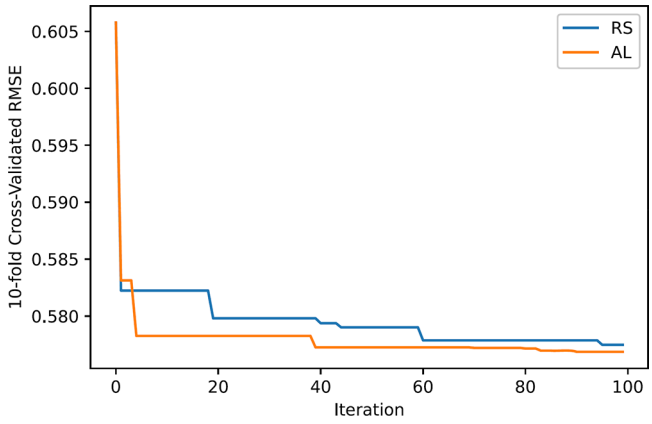


Figure 6. Convergence of the hyperparameters to the optimum state for the SVM model. AL has shown a better performance.

Table 6. Optimized Values Acquired by Different Methods for the Developed SVM Model

	range	RS	AL
C	0–10	9.049678	5.276798
gamma	0–1	0.195621	0.27791
best error		0.5775	0.5769

that given computational task. The learning procedures are performed by experts presented by neural networks, and the combination of which leads to the building of a Committee Machine (CM). Thus, solutions obtained from each expert are grouped and evidently will dominate solutions from individual models.[161] Classification of committee machines occurs by their structures, namely, static and dynamic structures.[162]

The combination of networks is the main component in CM construction. Thereby, the method of combining should be carefully selected. One of these methods is linear averaging, where simple or weighted averaging can be used.[163] The weighted averaging is commonly used because the solutions are grouped per their accuracy.

The CMIS model intends to explore a superior solution using the already constructed models. An approach like weighted arithmetic averaging with a bias term was used in this regard. No restriction was applied to make the sum of coefficients (without bias term) equal to unity. In fact, it is similar to the LR method. The expert integration of CMIS can be expressed using the following equation:

$$Y^i_{\mathrm{CMIS}} = a_1 \times Y^i_{\mathrm{LR}} + a_2 \times Y^i_{\mathrm{MLP}} + a_3 \times Y^i_{\mathrm{SVM}} + a_0 \qquad (4)$$

The best developed model for each model family is utilized for the CMIS. Here, the original values were used not the scaled ones. The optimum coefficients were determined after 100 trials. The determined optimized values are shown in Table 7. The best 10-fold cross-validated RMSE value was determined to be 4.5091.

4. MODEL PERFORMANCE EVALUATION

In this section, different statistical error functions and graphical approaches along with the leverage method used to assess each model's accuracy are discussed.

4.1. Statistical Error Analysis. Different statistical approaches were adopted for model assessment in this study, including the coefficient of determination (R^2), root mean squared error (RMSE), standard deviation (SD), mean relative

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304–32326

Table 7. Optimum Coefficients for the Developed CMIS Models

	coefficients			
best error	a_1	a_2	a_3	a_0
4.5091	−0.0594932181	0.2786232151	0.7762859660	0.2130974135

Table 8. Statistical Parameters for the Developed Models[a]

model	subset	RMSE	MARD	MRD	R^2	SD	N
LR	training	13.143	9.687	1.658	0.2039	13.143	968
	testing	13.199	9.732	1.394	0.1975	13.198	242
	total	13.154	9.696	1.605	0.2024	13.154	1210
MLP	training	5.441	3.790	0.298	0.8636	5.441	968
	testing	8.293	5.818	0.102	0.7032	8.293	242
	total	6.119	4.195	0.259	0.8282	6.119	1210
SVM	training	4.750	2.797	0.478	0.8969	4.749	968
	testing	8.070	5.567	0.236	0.7044	8.067	242
	total	5.575	3.351	0.430	0.8569	5.574	1210
CMIS	training	**4.622**	**2.810**	**0.247**	**0.9015**	**4.622**	968
	testing	**7.757**	**5.267**	**0.072**	**0.7307**	**7.755**	242
	total	**5.397**	**3.301**	**0.212**	**0.8661**	**5.397**	1210

[a]The best values for the total dataset are in bold. The CMIS has yielded the best performance considering all the parameters.

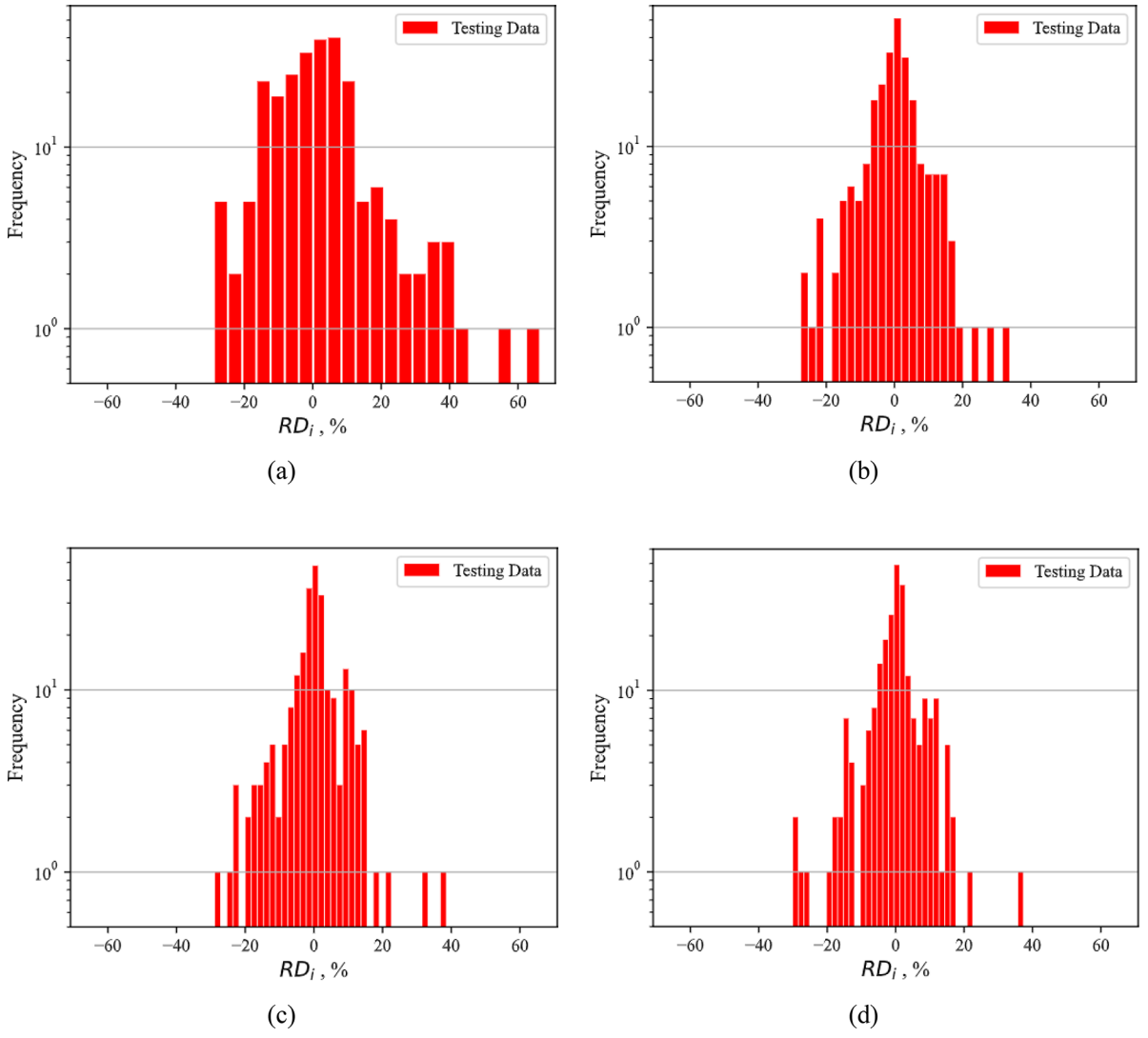


Figure 7. Relative deviation distribution plot for (a) LR, (b) MLP, (c) SVM, and (d) CMIS. The distribution for the CMIS model is closer to the normality compare with the other models.

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

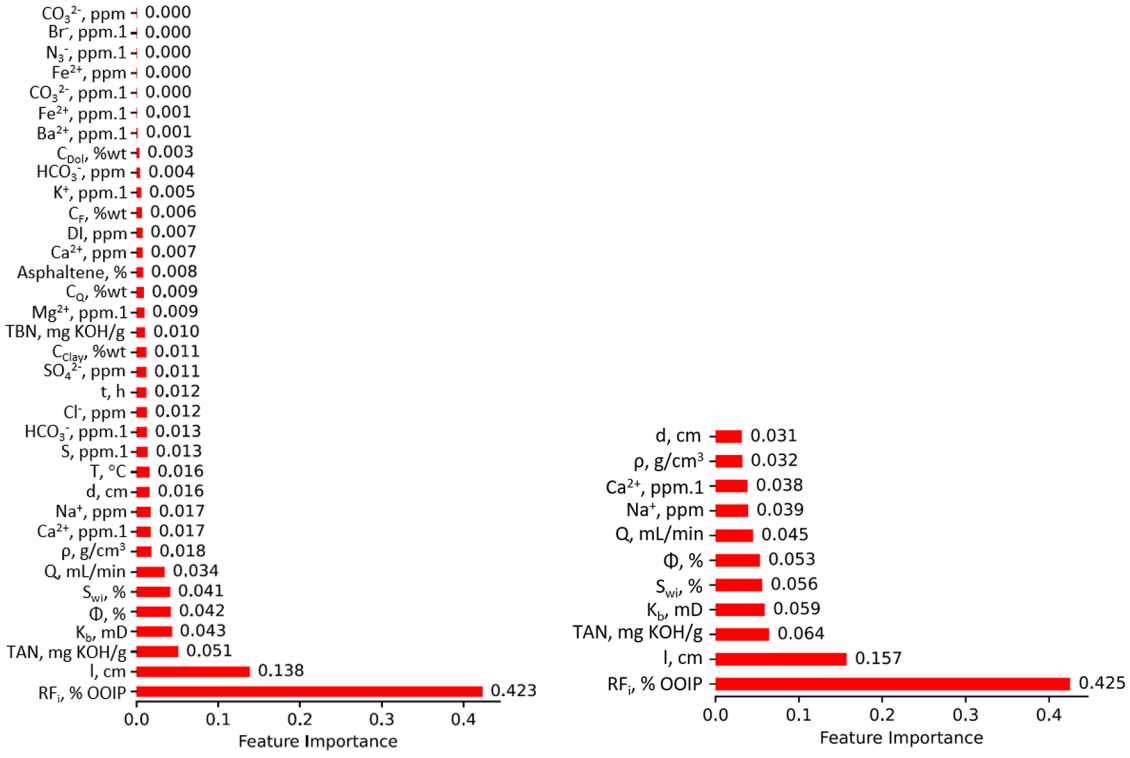


Figure 8. Features' importance of a DT-based model developed by (a) 35 and (b) 11 independent features. As shown here, the top 11 important features shown in the subplots are not identical.

deviation (MRD), and mean absolute relative deviation (MARD) were calculated for each model. The corresponding equations are shown below:

$$D_i = X^i_{\text{Pred.}} - X^i_{\text{Exp.}} \tag{5}$$

$$\text{RD}_i = \frac{D_i}{|X^i_{\text{Pred.}}| + |X^i_{\text{Exp.}}|} \times 100 \tag{6}$$

$$\text{MRD} = \frac{1}{N}\sum_{i=1}^{N} \text{RD}_i \tag{7}$$

$$\text{ARD}_i = |\text{RD}_i| \tag{8}$$

$$\text{MARD} = \frac{1}{N}\sum_{i=1}^{N} \frac{D_i}{|X^i_{\text{Pred.}}| + |X^i_{\text{Exp.}}|} \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} D_i^2}{\sum_{i=1}^{N} (X^i_{\text{Pred.}} - \bar{X}_{\text{Exp.}})^2}, \quad \bar{X}_{\text{Exp.}} = \frac{1}{N}\sum_{i=1}^{N} X^i_{\text{Exp.}} \tag{10}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (D_i)^2} \tag{11}$$

$$\text{SD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (D_i - \bar{D}_i)^2} \tag{12}$$

Each dataset is statistically analyzed. The best accuracy values for the testing dataset in Table 8 are in bold. Considering all of the statistical parameters, the CMIS model has yielded the best results. The accuracy of the developed models is in the order of CMIS > SVM > MLP > LR. Although the CMIS has shown superior statistical error values, its computational cost is much higher than the individual models. Considering this fact, the SVM can be selected as the best model. The developed models in this study are all data-based. A general drawback with such models is that they might not be able to consider the physics of the process. This limitation can be mitigated by using a larger database. For this study, all the available reported data were gathered for model development. The other limitation with the MLP and SVM models is the fact that they are black boxes. The developed models are provided in a user-friendly interface as the Supporting Information. They receive the input parameters in an excel file and deliver the prediction in another excel file. The instruction on how to use the software is provided in the Supporting Information.

4.2. Graphical Error Analysis. Herein, the RD_i distributions for all of the constructed models are compared (Figure 7). Herein, Figure 7a presents the distribution for the LR model, (b) for the MLP, (c) for the SVM, and (d) for the CMIS model. A roughly symmetrical distribution around zero is observed as one can see in Figure 7. The LR shows one extreme RD_i value of more than 60%. The positive value denotes an overprediction. The CMIS has the best symmetry with an MRD value of 0.072%. As is clear, the CMIS has the most compact plot around zero.

4.3. Leverage Method. To estimate the model reliability, its applicability domain must be identified. Applicability domain refers to the determination of outliers, which is a datum, or group of data. Outliers differ from the bulk of data and thereby should be eliminated. The leverage approach is a commonly used method for outlier determination.[164] The

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

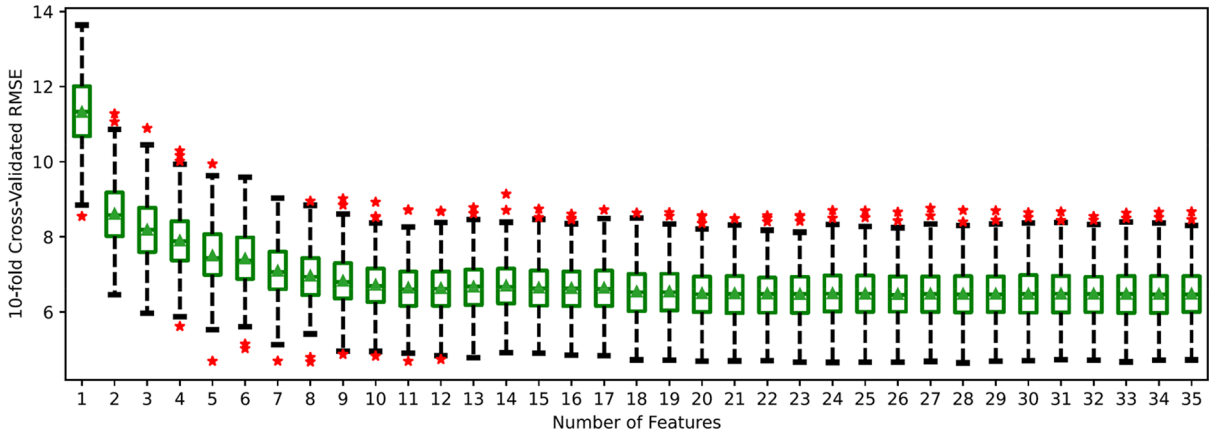


Figure 9. RMSE values for the models developed by different numbers of features. After passing the 11 features, the RMSE sharply increases.

Table 9. Selected Independent Features for Modeling[a]

	Operational Parameters	Rock Properties	Oil Properties	Brine Properties	Connate Water Properties
1	~~T, °C~~	l, cm	ρ, g/cm^3	~~HCO_3^-, ppm~~	~~HCO_3^-, ppm~~
2	Q, mL/min	d, cm	~~μ, cP~~	~~Cl^-, ppm~~	~~Cl^-, ppm~~
3	~~t, h~~	~~Φ, %~~	TAN, mg KOH/g	~~BO_2^-, ppm~~	~~BO_2^-, ppm~~
4	RF_i, % OOIP	~~K_h, mD~~	TBN, mg KOH/g	~~F^-, ppm~~	~~F^-, ppm~~
5	~~Flooding Stage~~	~~S_{wi}, %~~	~~*Saturate*, %~~	~~NO_3^-, ppm~~	~~NO_3^-, ppm~~
6		~~C_{Clay}, %wt~~	~~*Aromatic*, %~~	~~Br^-, ppm~~	~~Br^-, ppm~~
7		~~C_Q, %wt~~	~~*Resin*, %~~	~~N_3^-, ppm~~	~~N_3^-, ppm~~
8		~~C_{Mic}, %wt~~	~~*Asphaltene*, %~~	Na^+, ppm	~~Na^+, ppm~~
9		~~C_{Cal}, %wt~~		~~K^+, ppm~~	~~K^+, ppm~~
10		~~C_{Dol}, %wt~~		~~Li^+, ppm~~	~~Li^+, ppm~~
11		~~C_{Car}, %wt~~		~~SO_4^{2-}, ppm~~	~~SO_4^{2-}, ppm~~
12		~~C_A, %wt~~		~~Mg^{2+}, ppm~~	~~Mg^{2+}, ppm~~
13		~~C_F, %wt~~		~~Ca^{2+}, ppm~~	Ca^{2+}, ppm
14				~~Ba^{2+}, ppm~~	~~Ba^{2+}, ppm~~
15				~~Sr^{2+}, ppm~~	~~Sr^{2+}, ppm~~
16				~~CO_3^{2-}, ppm~~	~~CO_3^{2-}, ppm~~
17				~~Fe^{2+}, ppm~~	~~Fe^{2+}, ppm~~
18				~~$S_2O_4^{2-}$, ppm~~	~~$S_2O_4^{2-}$, ppm~~
19				~~Fe^{3+}, ppm~~	~~Fe^{3+}, ppm~~
20				~~*MI*, ppm~~	~~*MI*, ppm~~
21				~~*DI*, ppm~~	~~*DI*, ppm~~
22				~~*S*, ppm~~	~~*S*, ppm~~

[a]The excluded variables during the preprocessing stage are crossed out.

leverages or hat values (h_i) are estimated by using the Hat matrix (H), which is given in eq 13:

$$H = X(X^TX)^{-1}X^T \tag{13}$$

where H is the Hat matrix, X is an $l \times m$ matrix, in which l refers to the data points and m denotes the input parameters, and X^T indicates the transpose matrix X. The diagonal elements of a given Hat matrix are known as h_i.

The threshold leverage, known as the warning leverage (h^*) above which data is reported as distant, is determined by

$$h^* = 3 \times \frac{(N+1)}{b} \tag{14}$$

where N is the number of model parameters and b is the number of data used for model training.

5. RESULTS AND DISCUSSION

In this section, first, the feature selection followed by models' accuracy, flow assurance, and EOR implications is presented and discussed.

5.1. Feature Selection for Modeling. Overfitting is a common problem in AI modeling practices associated with a large number of independent parameters.[165] Furthermore, irrelevant features impose unfavorable influences on the performance of a wide range of models.[166] These facts dictate the importance of recognizing relevant features before

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

modeling, through a process called feature selection. Some predictive models such as the DT give the features' importance after the modeling was developed. In this way, it is possible to sort features based on their contribution to the modeling. At this point, the least important parameter is excluded and the modeling process is repeated again. This process can be continued until the desired number of features remains. The process of removing the least important feature in a stepwise way is called recursive feature elimination (RFE). Features' importance for the model developed by the whole 35 features is shown in Figure 8. It also shows that the top 11 important features shown in the subplots are not identical. This implies that the RFE method has considered the effect of independent parameters on each other.

The box plot for the 10-fold cross-validated models developed by the different number of input features is shown in Figure 9. The vertical axis shows the RMSE value. This plot indicates the existence of several irrelevant features. Although the best mean RMSE value is acquired for the 23 features, dropping more features does not considerably affect the RMSE. In this research work, 11 parameters were chosen for modeling since removing more features results in a relatively large jump in RMSE. Table 9 lists the selected and dropped features. All five categories have at least one representative. The important features selected by our model are in line with many of the experimental works reported in the literature.[17,44,167,168] Jadhunandan[167] and Al-Nofli et al.[169] proved the important role of the initial water saturation as well as brine composition in the wetting state of the rocks. They observed that a higher concentration of Ca^{2+} cations along with lower initial water saturations reduces the water wetness. Yang et al.[44] observed that waterflooding with NaCl brine improved the oil recovery, while with $CaCl_2$ brine with similar ionic strength, the improvement in oil recovery was not observed. This observation is explained based on the strong linkage between calcium ions and the rock surface coupled with the lack of capability for these cations to replace polar oil components on the rock surface. Austad et al.[17] observed more oil recovery with crude oils with higher z and low TBN compared to crude oils with low TAN and high TBN values. This is so since the basic components of crude oil are also able to attach to the silica surface and render its wettability condition.

5.2. Model Accuracy. As previously mentioned in this text, the CMIS has yielded the best performance. Hence, additional graphical accuracy analyses are provided only for the CMIS. The cross plot in which the vertical and horizontal axes are predicted and experimental RF_f, respectively, is shown in Figure 10. As one can see in Figure 10, the majority of the data points are concentrated in the close vicinity of the bisector line for both training and testing datasets. In addition, most of the points are accumulated between 30 and 90%. This was already shown by the box plot (Figure 4).

The RD_i based on the RF_f is shown in Figure 11. The maximum and minimum RD_i values are about 60 and −40, respectively. However, most of the RD_i values are bounded in a range of −20 to 20.

Simultaneous representation of experimental and predicted RF_f values versus the data point index is shown in Figure 12. The purple diamonds and blue solid lines are the experimental and predicted values, respectively. This figure shows that the model forecasting with very few exceptions interconnects most of the experimental data.

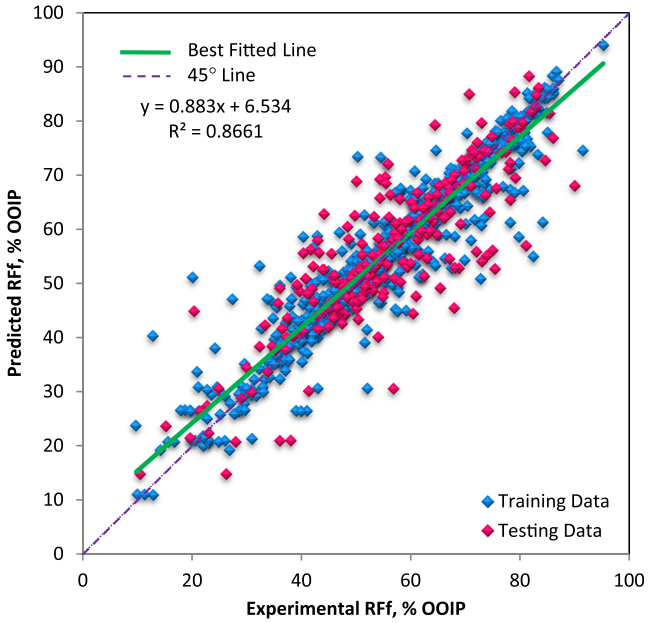


Figure 10. Cross plot of the predicted and experimental RF_f values for training and testing dataset for the CMIS. Accumulation of majority of the data points around the bisector line shows performance of the model.

ARD_i distribution for the CMIS model over the two-dimensional region of the most influential parameter (RF_i, % OOIP) and the other input parameters are shown in Figure 13. As one can see here, the majority of ARD_i values are less than 2% for the upper half of the RF_i and % OOIP values. For the same input variable, the greatest values of ARD_i have happened for values less than 20%. As indicated on the subplot (h), the model has shown poor performance for Ca^{2+} (ppm) in the close vicinity of 20,000 ppm for which the ARD_i is as high as about 47%.

A visualized cumulative frequency of ARD_i (%) for the developed LR, MLP, SVM, and CMIS models is presented in Figure 14. The cumulative frequency is plotted against the ARD_i (%). As one can see, the CMIS, SVM, and MLP have performed much better than the LR. In other words, this plot compares the models with portions of the database predicted with absolute errors less than a typical ARD_i. For example, the CMIS, SVM, MLP, and LR models have predicted, respectively, about 93, 91, 90, and 63% of the dataset with the ARD_i values of $\leq$10%.

5.3. Outlier Diagnosis. William's plot is used to analyze the measurement outliers for the CMIS model (Figure 15). The h^* and residual limits bound the safe zone, which is bounded by the $SD_i = \pm 3$ (horizontal green lines) and h^* (purple vertical line). As evident, most of the data points are located in this zone. One reason for the absence of the high-leverage data is excluding the interesting outliers in the preprocessing step. As shown in Figure 15, there are 15 and 19 outliers for the training and testing datasets, respectively. The 34 outliers are listed in Table 10 with their SD_i and h_i.

5.4. EOR Implications. Providing reliable and accurate predictive AI models can help with better performance production, process design, and optimization concerning LSWF in sandstone reservoirs. In addition, the database assembled during the process of this research work can be used

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

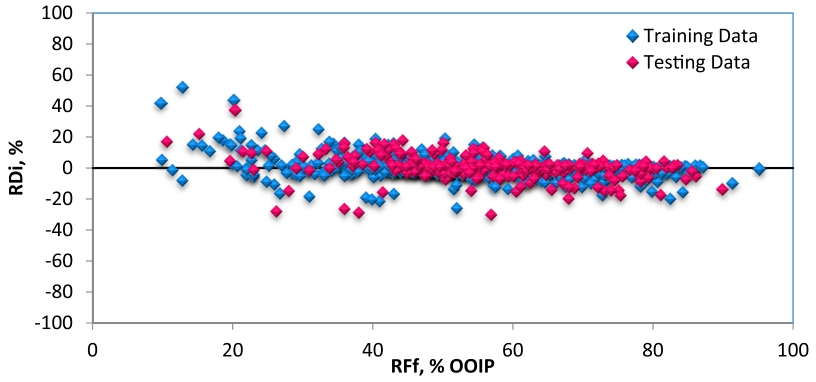
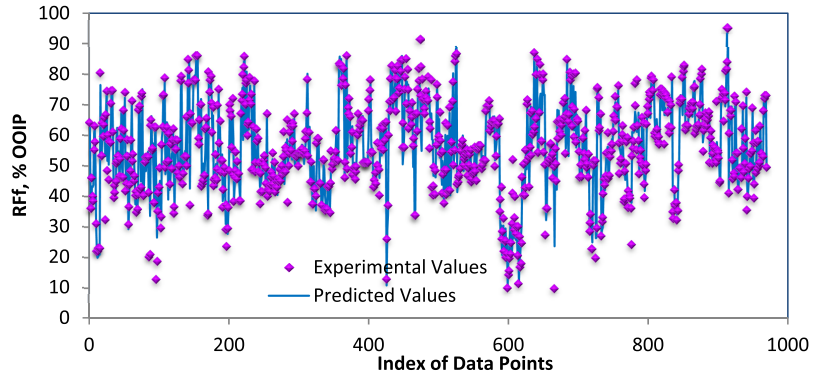
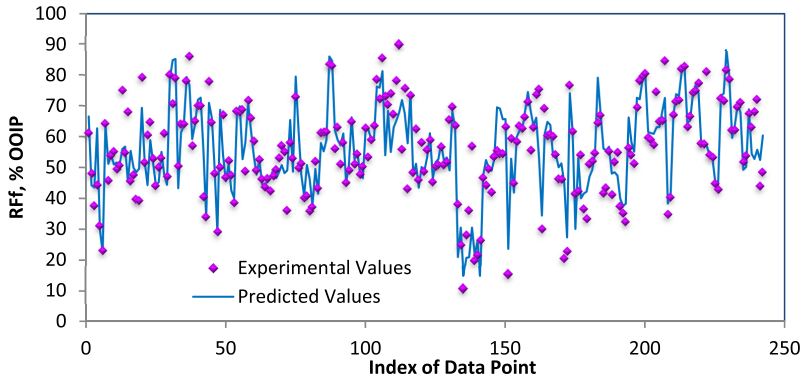


Figure 11. Relative deviation plot for training and testing data. The more accumulation of data points close to the zero line, the more accurate the model.



(a)



(b)

Figure 12. Simultaneous depiction of predicted and experimental values based on the index of data points for (a) training and (b) testing datasets. As is illustrated, the CMIS model follows the experimental values.

by researchers to further our understanding about mechanisms behind LSWF in sandstone reservoirs.

6. CONCLUSIONS

In this research work, first, a comprehensive database of 1316 data points was assembled from 99 experimental studies on LSWF in sandstone reported in the literature. This is so far the largest database ever collected or reported. Then, the database was analyzed for removing problematic samples and recognizing the most relevant input variables. For this, several statistical operations including processes such as duplicate and low-variance feature removal, missing value imputation, collinearity assessment, outlier removal, and feature selection for modeling were performed. In the next step, four different predictive models of LR, MLP, SVM, and CMIS were developed and their performance and applicability domain were analyzed using different statistical and graphical methods. The main finding of this research work can be summarized as follows:

1. Analysis of the experimental data collected from the literature revealed that the influential parameters on the LSWF performance could be categorized into five general categories of rock, oil, brine, connate water properties, and operational parameters.
2. Using a recursive feature elimination, the most relevant parameters were recognized as Q (mL/min), RF_i (%

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

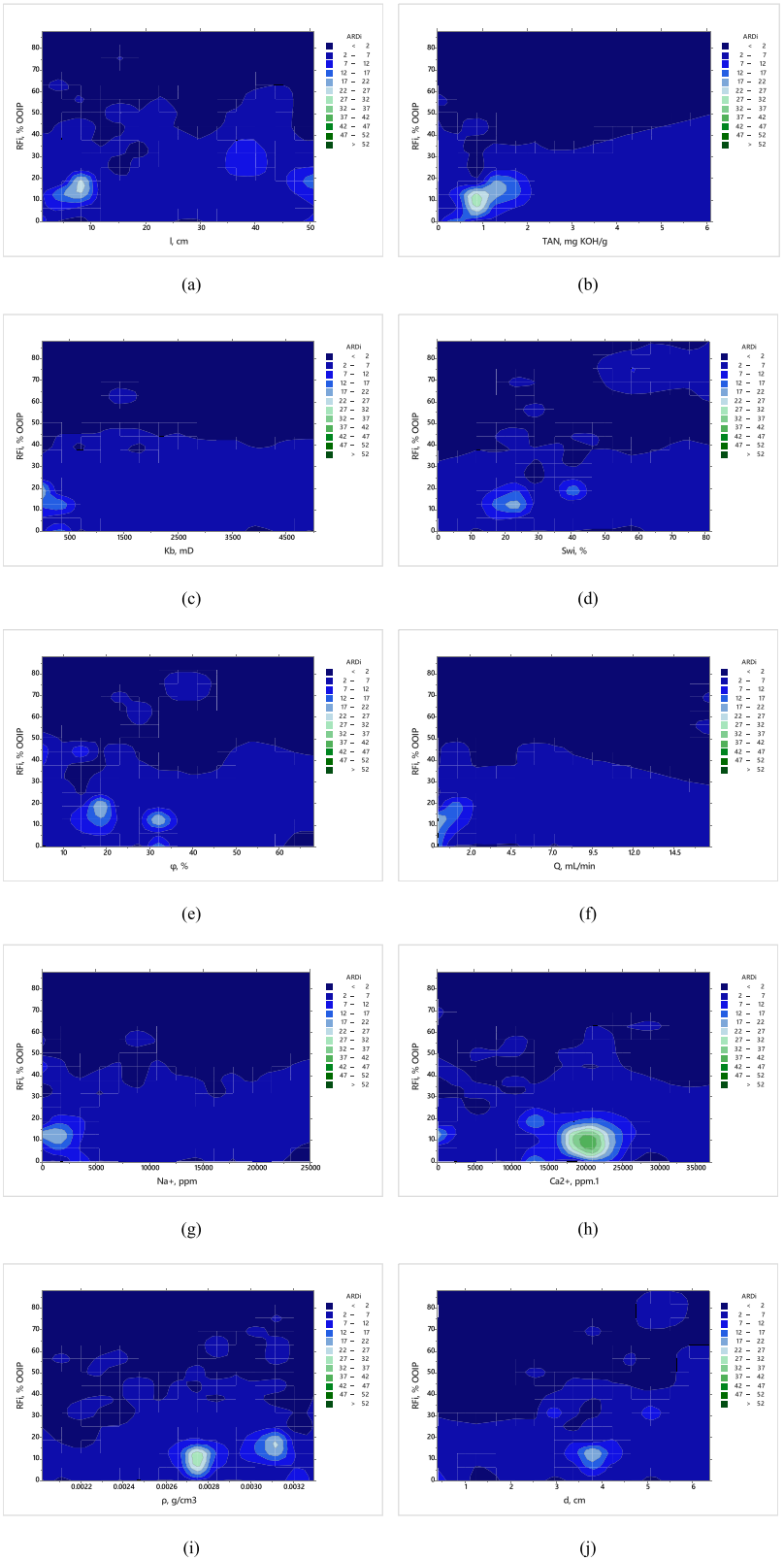


Figure 13. Contour map analysis of the prediction error for the CMIS model developed in this research work for (a) RF_i (% OOIP) versus l (cm), (b) RF_i (% OOIP) versus TAN (mg KOH g), (c) RF_i (% OOIP) versus K_b (mD), (d) RF_i (% OOIP) versus S_{wi} (%), (e) RF_i (% OOIP) versus Φ (%), (f) RF_i (% OOIP) versus Q (mL/min), (g) RF_i (% OOIP) versus Na^+ (ppm), (h) RF_i (% OOIP) versus Ca^{2+} (ppm), (i) RF_i (% OOIP) versus ρ (g/cm^3), and (j) RF_i (% OOIP) versus d (cm).

OOIP), l (cm), d (cm), Φ (%), K_b (mD), S_{wi} (%), ρ (g/cm^3), Na^+ (ppm), and Ca^{2+} (ppm). These statistically determined important features represent all of the five categories of independent variables and were in line with numerous experimental works reported in the literature.

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

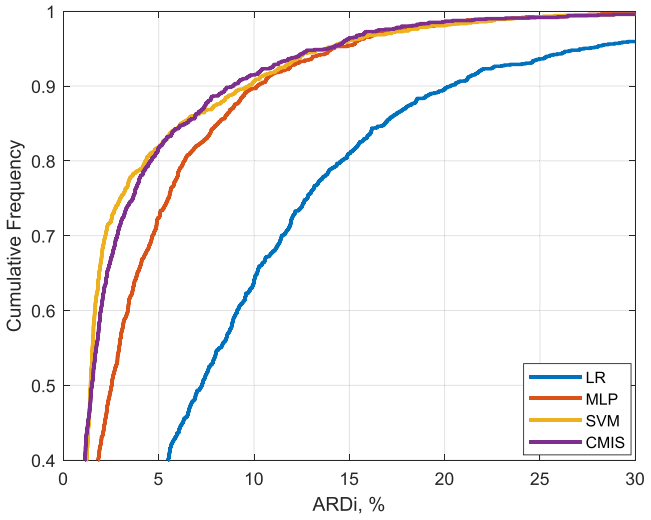


Figure 14. Cumulative frequency of error for the four models developed in this research work. The best models are in the order of CMIS > SVM > MLP > LR.

3. Among the developed models, the CMIS showed the best performance by RMSE values of 4.622 and 7.757 for the training and testing datasets, respectively. The corresponding MARD values were 2.810 and 5.267%, respectively. Additionally, a mean relative deviation of 0.072% was demonstrated symmetrical error distribution for this model.
4. Analysis of the applicability domain determined 34 data samples as calculation outliers. This amount was equivalent to 2.8% of the database used. The outliers included 15 and 19 samples belonging to the training and testing datasets, respectively. In addition, five high-leverage samples were detected. This low amount of calculation outliers depicts the generality of the developed model.
5. The findings of this study can be used to further delineate mechanisms behind LSWF and optimization of this EOR process in sandstone reservoirs.

Table 10. Outlier Data Detected by the William's Plot

index	RF_f % OOIP	predicted RF_f % OOIP	h_i	SD_i	dataset	leverage status
1	32.3	53.169	0.004	3.875	training	low
2	72.68	50.782	0.004	−4.066	training	low
3	20.16	51.046	0.004	5.735	training	low
4	12.79	40.268	0.006	5.106	training	low
5	79.8	58.511	0.004	−3.954	training	low
6	91.4	74.459	0.005	−3.146	training	low
7	82.4	54.913	0.004	−5.104	training	low
8	55.7	72.019	0.005	3.031	training	low
9	40.4	58.500	0.004	3.361	training	low
10	84.09	61.204	0.004	−4.248	training	low
11	52	30.578	0.004	−3.978	training	low
12	54.42	73.158	0.035	3.534	training	low
13	50.34	73.281	0.035	4.327	training	low
14	27.34	47.007	0.004	3.652	training	low
15	78.1	61.155	0.004	−3.146	training	low
16	44.2	62.732	0.006	3.444	testing	low
17	75	56.019	0.009	−3.533	testing	low
18	67.93	45.390	0.004	−4.185	testing	low
19	50	68.748	0.009	3.489	testing	low
20	46.7	45.560	0.041	−0.216	testing	high
21	52.1	51.187	0.041	−0.173	testing	high
22	63	58.958	0.073	−0.778	testing	high
23	73	53.960	0.005	−3.537	testing	low
24	73.9	55.044	0.004	−3.501	testing	low
25	90	67.921	0.005	−4.102	testing	low
26	65.42	49.059	0.007	−3.041	testing	low
27	38	20.976	0.003	−3.160	testing	low
28	56.8	30.516	0.004	−4.881	testing	low
29	75.3	52.602	0.005	−4.217	testing	low
30	20.43	44.796	0.018	4.556	testing	low
31	57.5	58.180	0.052	0.129	testing	high
32	81	56.865	0.008	−4.490	testing	low
33	72	55.766	0.004	−3.014	testing	low
34	48.4	60.384	0.044	2.271	testing	high

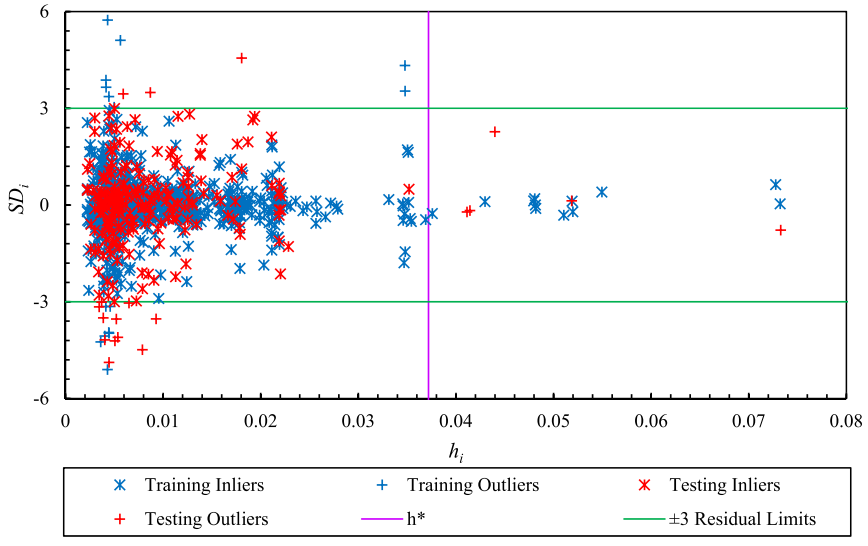


Figure 15. William's plot for the CMIS model developed in this research work. The great portion of the data points is rested in the safe zone bounded by ±3 *SDi* and *h**.

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.1c05493.

Developed models (ZIP)

Instructions on how to use the software (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Ali Shafiei − *Petroleum Engineering Program, School of Mining and Geosciences, Nazarbayev University, Nur−Sultan 010000, Kazakhstan;* orcid.org/0000-0002-7740-9109; Email: ali.shafiei@nu.edu.kz

Authors

Afshin Tatar − *Petroleum Engineering Program, School of Mining and Geosciences, Nazarbayev University, Nur−Sultan 010000, Kazakhstan*

Ingkar Askarova − *Petroleum Engineering Program, School of Mining and Geosciences, Nazarbayev University, Nur−Sultan 010000, Kazakhstan*

Mahsheed Rayhani − *Petroleum Engineering Program, School of Mining and Geosciences, Nazarbayev University, Nur−Sultan 010000, Kazakhstan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c05493

Author Contributions

A.S. contributed to conceptualization; A.T. and I.A. contributed to formal analysis, methodology, visualization, and validation; A.T., I.A., and M.R. contributed to data extraction; A.T., I.A., A.S., and M.R. contributed to writing (original draft preparation); A.T., I.A., A.S., and M.R. contributed to writing (review and editing); A.S. contributed to project administration, funding acquisition, resources, data curation, and supervision. All authors have read and agreed to the published version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to thank anonymous reviewers for their critical yet fair and constructive comments, which helped the authors to improve the quality of the manuscript. Financial support received from Nazarbayev University through a Faculty Development Competitive Research Grants Program (grant# 110119FD4529) is hereby acknowledged and highly appreciated.

■ NOMENCLATURE

Abbreviations

AD	applicability domain
AI	artificial intelligence
AL	annealling
ANFIS	adaptive neuro fuzzy inference system
ANN	artificial neural network
CM	committee machine
CMIS	committee machine intelligent system
CSA	coupled simulated annealing
DLE	double layer expansions
DNS	deep net simulator
DT	decision tree
EOR	enhanced oil recovery
ET	extra tree
FLS	fuzzy logic system
LR	linear regression
LSE	low salinity effect
LSSVM	least squares support vector machine
LSWF	low salinity waterflooding
ML	machine learning
MLP	multilayer perceptron
MLR	multiple linear regression
N.A.	not available
OOIP	original oil in place
RBF	radial basis function
RF	random forest
RF_f	final recovery factor
RF_i	initial recovery factor
RS	random search
SAGD	steam-assisted gravity drainage
SVM	support vector machine

Parameters

ARD_i	absolute relative deviation, %
Aromatic	aromatic content, %
Asphaltene	asphaltene content, %
b	number of data used for model training
Ba^{2+}	barium ion content, %
BO_2^-	metaborate ion content, %
Br^-	bromine ion content, %
C	regularization parameter for the SVM model
C_A	anhydrite content, %wt
Ca^{2+}	calcium ion content, %
C_{Cal}	calcite content, %wt
C_{Car}	carbonate content, %wt
C_{Clay}	clay content, %wt
C_{Dol}	dolomite content, %wt
C_F	feldspar content, %wt
Cl^-	chloride ion content, %
C_{Mic}	mica content, %wt
CO_3^{2-}	carbonate ion content, %
C_Q	quartz content, %wt
d	diameter of the core, cm
D_i	deviation, %
DI	total divalent ions, %
F^-	fluoride ion content, %
Fe^{2+}	ferrous ion content, %
Fe^{3+}	ferric ion content, %
gamma	kernel coefficient for the SVM model
H	Hat matrix
h^*	warning leverage
HCO_3^-	bicarbonate ion content, ppm
h_i	Hat values
IQR	interquartile range
k	Kurtosis value
K^+	potassium ion content, %
K_b	brine permeability, mD
l	length, cm
Li^+	lithium ion content, %
m	input parameters
MAPE	mean absolute percentage error
MARD	mean absolute relative deviation, %
max	maximum value
mean	mean value
median	median value

Mg^{2+}	magnesium ion content, %
MI	total monovalent ions, %
min	minimum value
mode	mode value
MRD	mean relative deviation
MSE	mean squared error
N	number of data points
N_3^-	azide ion content, %
Na^+	sodium ion content, %
NO_3^-	nitrate ion content, %
Q	injection flow rate, mL/min
Q_1	first quartile
Q_3	third quartile
R	Spearman correlation factor
R^2	coefficient of determination
range	range value
RD_i	relative deviation, %
Resin	resin content, %
RF_f	final oil recovery, % OOIP
RF_i	initial recovery factor, % OOIP
RMSE	root mean squared error, %
s	skewness value
S	salinity ion content, %
$S_2O_4^{2-}$	dithionite ion content, %
Saturate	saturate content, %
SD	standard deviation, %
SD_i	standardized deviation, %
SO_4^{2-}	sulfate ion content, %
Sr^{2+}	strontium ion content, %
S_{wi}	irreducible water saturation, %
T	temperature, °C
t	aging time, h
TAN	total acid number, mg KOH/g
TBN	total base number, mg KOH/g
Var	variance value
VIF	variance inflation factor
X^T	transpose matrix X
μ	viscosity, cP
ρ	density, g/cm^3
Φ	porosity, %

REFERENCES

(1) Alagorni, A. H.; Yaacob, Z. B.; Nour, A. H. An overview of oil production stages: enhanced oil recovery techniques and nitrogen injection. *Int. J. Environ. Sci. Dev.* **2015**, *6*, 693−701.

(2) Akhmetgareev, V.; Khisamov, R. 40 Years of Low-Salinity Waterflooding in Pervomaiskoye Field, Russia: Incremental Oil, *SPE European Formation Damage Conference and Exhibition*, Budapest, Hungary, 2015.

(3) Almeida da Costa, A.; Jaeger, P.; Santos, J.; Soares, J.; Trivedi, J.; Embiruçu, M.; Meyberg, G. The Influence of Rock Composition and pH on Reservoir Wettability for Low Salinity Water-CO2 EOR Applications in Brazilian Reservoirs, *SPE Annual Technical Conference and Exhibition*, Calgary, Alberta, Canada, 2019.

(4) Al-Otaibi, A. An investigation into the roles of chlorides and sulphate salts on the performance of low-salinity injection in sandstone reservoirs: experimental approach. *J. Pet. Explor. Prod. Technol.* **2020**, *10*, 2857−2871.

(5) Al-Saedi, H. N.; Qubian, A.; Al-Bazzaz, W.; Flori, R. Experimental Study of Low Salinity Water Flooding: The Effect of Polar Organic Components in Low-Permeable Sandstone Reservoir, *International Petroleum Technology Conference*, Dhahran, Kingdom of Saudi Arabia, 2020.

(6) Piñerez Torrijos, I. D.; Puntervold, T.; Strand, S.; Austad, T.; Bleivik, T. H.; Abdullah, H. I. An experimental study of the low salinity Smart Water - Polymer hybrid EOR effect in sandstone material. *J. Pet. Sci. Eng.* **2018**, *164*, 219−229.

(7) Mokheimer, E.; Hamdy, M.; Abubakar, Z.; Shakeel, M. R.; Habib, M. A.; Mahmoud, M. A Comprehensive Review of Thermal Enhanced Oil Recovery: Techniques Evaluation. *J. Energy Resour. Technol.* **2019**, *141* ().

(8) Jadhunandan, P. P.; Morrow, N. R. Effect of Wettability on Waterflood Recovery for Crude-Oil/Brine/Rock Systems. *SPE Reservoir Eng.* **1995**, *10*, 40−46.

(9) Awolayo, A. N.; Sarma, H. K.; Nghiem, L. X. Brine-dependent recovery processes in carbonate and sandstone petroleum reservoirs: review of laboratory-field studies, interfacial mechanisms and modeling attempts. *Energies* **2018**, *11*, 3020.

(10) Morrow, N. R. Wettability and Its Effect on Oil Recovery. *J. Pet. Technol.* **1990**, *42*, 1476−1484.

(11) Morrow, N. R.; Tang, G. Q.; Valat, M.; Xie, X. Prospects of improved oil recovery related to wettability and brine composition. *J. Pet. Sci. Eng.* **1998**, *20*, 267−276.

(12) Tang, G. Q.; Morrow, N. R. Salinity, Temperature, Oil Composition, and Oil Recovery by Waterflooding. *SPE Reservoir Eng.* **1997**, *12*, 269−276.

(13) Tang, G.-Q.; Morrow, N. R. Influence of brine composition and fines migration on crude oil/brine/rock interactions and oil recovery. *J. Pet. Sci. Eng.* **1999**, *24*, 99−111.

(14) Kim, Y.; Kim, C.; Kim, J.; Kim, Y.; Lee, J. Experimental investigation on the complex chemical reactions between clay minerals and brine in low salinity water-flooding. *J. Ind. Eng. Chem.* **2020**, *89*, 316−333.

(15) Mahzari, P.; Sohrabi, M. Crude Oil/Brine Interactions and Spontaneous Formation of Micro-Dispersions in Low Salinity Water Injection, *SPE Improved Oil Recovery Symposium*, Tulsa, Oklahoma, 2014.

(16) Nasralla, R. A.; Nasr-El-Din, H. A. Double-layer expansion: is it a primary mechanism of improved oil recovery by low-salinity waterflooding? *SPE Reservoir Eval. Eng.* **2014**, *17*, 49−59.

(17) Austad, T.; RezaeiDoust, A.; Puntervold, T. Chemical Mechanism of Low Salinity Water Flooding in Sandstone Reservoirs, *SPE Improved Oil Recovery Symposium*, Tulsa, Oklahoma, USA, 2010.

(18) Lager, A.; Webb, K. J.; Black, C. J. J.; Singleton, M.; Sorbie, K. S. Low Salinity Oil Recovery - An Experimental Investigation1. *Petrophys.* **2008**, *49*, 28−35.

(19) Rayhani, M.; Simjoo, M.; Chahardowli, M. Insights into effects of water chemistry on the sandstone wetting characteristics. *J. Pet. Sci. Eng.* **2020**, *195*, 107929.

(20) Wang, L.; Fu, X. Data-driven analyses of low salinity water flooding in sandstones. *Fuel* **2018**, *234*, 674−686.

(21) Nasralla, R. A.; Bataweel, M. A.; Nasr-El-Din, H. A. Investigation of Wettability Alteration and Oil-Recovery Improvement by Low-Salinity Water in Sandstone Rock. *J. Can. Pet. Technol.* **2013**, *52*, 144−154.

(22) Ligthelm, D. J.; Gronsveld, J.; Hofman, J.; Brussee, N.; Marcelis, F.; van der Linde, H. Novel Waterflooding Strategy By Manipulation Of Injection Brine Composition, *EUROPEC/EAGE Conference and Exhibition*, Amsterdam, The Netherlands, 2009.

(23) Mahani, H.; Berg, S.; Ilic, D.; Bartels, W. B.; Joekar-Niasar, V. Kinetics of Low-Salinity-Flooding Effect. *SPE J.* **2015**, *20*, 8−20.

(24) Fredriksen, S. B.; Rognmo, A. U.; Sandengen, K.; Fernø, M. A. Wettability Effects on Osmosis as an Oil-Mobilization Mechanism During Low-Salinity Waterflooding. *Petrophys.* **2017**, *58*, 28−35.

(25) Emadi, A.; Sohrabi, M. Visual Investigation of Oil Recovery by Low Salinity Water Injection: Formation of Water Micro-Dispersions and Wettability Alteration, *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, 2013.

(26) Alvarado, V.; Bidhendi, M. M.; Garcia-Olvera, G.; Morin, B.; Oakey, J. S. Interfacial Visco-Elasticity of Crude Oil - Brine: An Alternative EOR Mechanism in Smart Waterflooding, *SPE Improved Oil Recovery Symposium*, Tulsa, Oklahoma, USA, 2014.

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

(27) Katende, A.; Sagala, F. A critical review of low salinity water flooding: Mechanism, laboratory and field application. *J. Mol. Liq.* **2019**, *278*, 627−649.
(28) Tabrizy, V. A.; Hamouda, A. A.; Denoyel, R. Influence of Magnesium and Sulfate Ions on Wettability Alteration of Calcite, Quartz, and Kaolinite: Surface Energy Analysis. *Energy Fuels* **2011**, *25*, 1667−1680.
(29) Xie, Q.; Liu, F.; Chen, Y.; Yang, H.; Saeedi, A.; Hossain, M. M. Effect of electrical double layer and ion exchange on low salinity EOR in a pH controlled system. *J. Pet. Sci. Eng.* **2019**, *174*, 418−424.
(30) Xie, Q.; Liu, Y.; Wu, J.; Liu, Q. Ions tuning water flooding experiments and interpretation by thermodynamics of wettability. *J. Pet. Sci. Eng.* **2014**, *124*, 350−358.
(31) Rivet, S. M.; Lake, L. W.; Pope, G. A. A Coreflood Investigation of Low-Salinity Enhanced Oil Recovery, *SPE Annual Technical Conference and Exhibition*, Florence, Italy, 2010.
(32) You, Z.; Badalyan, A.; Yang, Y.; Bedrikovetsky, P.; Hand, M. Fines migration in geothermal reservoirs: Laboratory and mathematical modelling. *Geothermics* **2019**, *77*, 344−367.
(33) Myint, P. C.; Firoozabadi, A. Thin liquid films in improved oil recovery from low-salinity brine. *Curr. Opin. Colloid Interface Sci.* **2015**, *20*, 105−114.
(34) Dang, C. T.; Nghiem, L. X.; Chen, Z.; Nguyen, Q. P. Modeling Low Salinity Waterflooding: Ion Exchange, Geochemistry and Wettability Alteration, *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, 2013.
(35) Piñerez, T. I. D.; Austad, T.; Strand, S.; Puntervold, T.; Wrobel, S.; Hamon, G. Linking Low Salinity EOR Effects in Sandstone to pH, Mineral Properties and Water Composition, *SPE Improved Oil Recovery Conference*, Tulsa, Oklahoma, USA, 2016; Tulsa, Oklahoma, USA, 2016.
(36) Aksulu, H.; Håmsø, D.; Strand, S.; Puntervold, T.; Austad, T. Evaluation of Low-Salinity Enhanced Oil Recovery Effects in Sandstone: Effects of the Temperature and pH Gradient. *Energy Fuels* **2012**, *26*, 3497−3503.
(37) Afekare, D. A.; Radonjic, M. From Mineral Surfaces and Coreflood Experiments to Reservoir Implementations: Comprehensive Review of Low-Salinity Water Flooding (LSWF). *Energy Fuels* **2017**, *31*, 13043−13062.
(38) Shehata, A. M.; Nasr-El-Din, H. A. Role of Sandstone Mineral Compositions and Rock Quality on the Performance of Low-Salinity Waterflooding, *International Petroleum Technology Conference*, Kuala Lumpur, Malaysia, 2014.
(39) Wickramathilaka, S.; Howard, J.; Morrow, N.; Buckley, J. An experimental study of low salinity waterflooding and spontaneous imbibition, *IOR 2011-16th European Symposium on Improved Oil Recovery*, 2011; European Association of Geoscientists & Engineers: 2011; pp cp-230-00051.
(40) Zekri, A. Y.; Nasr, M. S.; Al-Arabai, Z. I. Effect of LoSal on Wettability and Oil Recovery of Carbonate and Sandstone Formation, *International Petroleum Technology Conference*, Bangkok, Thailand, 2011.
(41) Bartels, W.-B.-B.; Mahani, H.; Berg, S.; Menezes, R.; van der Hoeven, J. A.; Fadili, A. Oil Configuration Under High-Salinity and Low-Salinity Conditions at Pore Scale: A Parametric Investigation by Use of a Single-Channel Micromodel. *SPE J.* **2017**, *22*, 1362−1373.
(42) Al-Saedi, H. N.; Brady, P. V.; Flori, R.; Heidari, P. Novel Insights into Low Salinity Water Flooding Enhanced Oil Recovery in Sandstone: The Clay Role Study, *SPE Improved Oil Recovery Conference*, Tulsa, Oklahoma, USA, 2018.
(43) Nasralla, R. A.; Nasr-El-Din, H. A. Impact of cation type and concentration in injected brine on oil recovery in sandstone reservoirs. *J. Pet. Sci. Eng.* **2014**, *122*, 384−395.
(44) Yang, J.; Dong, Z.; Lin, M. The Impact of Brine Composition and Salinity on the Wettability of Sandstone. *Pet. Sci. Technol.* **2015**, *33*, 430−436.
(45) Miyauchi, T. E.; Lu, Y.; Firoozabadi, A. Low Salinity Water Injection: Effect of Acid and Base Functionality on Recovery Performance, *SPE Annual Technical Conference and Exhibition*, San Antonio, Texas, USA, 2017.
(46) Garcia-Olvera, G.; Reilly, T. M.; Lehmann, T. E.; Alvarado, V. Effects of asphaltenes and organic acids on crude oil-brine interfacial visco-elasticity and oil recovery in low-salinity waterflooding. *Fuel* **2016**, *185*, 151−163.
(47) McMillan, M. D.; Rahnema, H.; Romiluy, J.; Kitty, F. J. Effect of exposure time and crude oil composition on low-salinity water flooding. *Fuel* **2016**, *185*, 263−272.
(48) RezaeiDoust, A.; Puntervold, T.; Austad, T. Chemical Verification of the EOR Mechanism by Using Low Saline/Smart Water in Sandstone. *Energy Fuels* **2011**, *25*, 2151−2162.
(49) Fjelde, I.; Omekeh, A. V.; Sokama-Neuyam, Y. A. Low Salinity Water Flooding: Effect Of Crude Oil Composition, *SPE Improved Oil Recovery Symposium*, Tulsa, Oklahoma, USA, 2014; Tulsa, Oklahoma, USA, 2014.
(50) Irannezhad, A.; Primkulov, B.; Juanes, R.; Zhao, B. Fluid-fluid displacement in mixed-wet porous media. *AGU Fall Meeting Abstracts* **2020**, *2020*, H018−H005.
(51) Ivuawuogu, H.; Cheng, Y.; Zhang, Y.; Khataniar, S. Experimental Investigation of the Role of Different Clays in Low Salinity Waterflooding, *SPE Annual Technical Conference and Exhibition*, Virtual, 2020.
(52) Kakati, A.; Sangwai, J. S. Wettability Alteration of Mineral Surface during Low-Salinity Water Flooding: Role of Salt Type, Pure Alkanes, and Model Oils Containing Polar Components. *Energy Fuels* **2018**, *32*, 3127−3137.
(53) Mahmoudvand, M.; Javadi, A.; Pourafshary, P.; Vatanparast, H.; Bahramian, A. Effects of cation salinity on the dynamic interfacial tension and viscoelasticity of a water-oil system. *J. Pet. Sci. Eng.* **2021**, *206*, 108970.
(54) Mamonov, A.; Puntervold, T.; Strand, S.; Hetland, B.; Andersen, Y.; Wealth, A.; Nadeau, P. H. Contribution of Feldspar Minerals to pH during Smart Water EOR Processes in Sandstones. *Energy Fuels* **2020**, *34*, 55−64.
(55) Esene, C.; Zendehboudi, S.; Shiri, H.; Aborig, A. Deterministic tools to predict recovery performance of carbonated water injection. *J. Mol. Liq.* **2020**, *301*, 111911.
(56) Moosavi, S. R.; Wood, D. A.; Ahmadi, M. A.; Choubineh, A. ANN-Based Prediction of Laboratory-Scale Performance of CO2-Foam Flooding for Improving Oil Recovery. *Nat. Resour. Res.* **2019**, *28*, 1619−1637.
(57) Zendehboudi, S.; Rezaei, N.; Lohi, A. Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. *Appl. Energy* **2018**, *228*, 2539−2566.
(58) Ghiasi, M. M.; Mohammadi, A. H. Application of decision tree learning in modelling CO2 equilibrium absorption in ionic liquids. *J. Mol. Liq.* **2017**, *242*, 594−605.
(59) Ghiasi, M. M.; Yarveicy, H.; Arabloo, M.; Mohammadi, A. H.; Behbahani, R. M. Modeling of stability conditions of natural gas clathrate hydrates using least squares support vector machine approach. *J. Mol. Liq.* **2016**, *223*, 1081−1092.
(60) Tatar, A.; Shokrollahi, A.; Mesbah, M.; Rashid, S.; Arabloo, M.; Bahadori, A. Implementing Radial Basis Function Networks for modeling CO2-reservoir oil minimum miscibility pressure. *J. Nat. Gas Sci. Eng.* **2013**, *15*, 82−92.
(61) Shokrollahi, A.; Tatar, A.; Safari, H. On accurate determination of PVT properties in crude oil systems: Committee machine intelligent system modeling approach. *J. Taiwan Inst. Chem. Eng.* **2015**, *55*, 17−26.
(62) Zheng, J.; Leung, J. Y.; Sawatzky, R. P.; Alvarez, J. M. An AI-based workflow for estimating shale barrier configurations from SAGD production histories. *Neural Comput. Appl.* **2019**, *31*, 5273−5297.
(63) Tariq, Z.; Mahmoud, M.; Abdulraheem, A. An intelligent data-driven model for Dean−Stark water saturation prediction in carbonate rocks. *Neural Comput. Appl.* **2020**, *32*, 11919−11935.
(64) Hassan, A.; Aljawad, M. S.; Mahmoud, M. An Artificial Intelligence-Based Model for Performance Prediction of Acid

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

Fracturing in Naturally Fractured Reservoirs. *ACS Omega* **2021**, *6*, 13654−13670.
(65) Desouky, M.; Tariq, Z.; Aljawad, M. S.; Alhoori, H.; Mahmoud, M.; Abdulraheem, A. Machine Learning-Based Propped Fracture Conductivity Correlations of Several Shale Formations. *ACS Omega* **2021**, *6*, 18782−18792.
(66) Kalam, S.; Abu-Khamsin, S. A.; Al-Yousef, H. Y.; Gajbhiye, R. A novel empirical correlation for waterflooding performance prediction in stratified reservoirs using artificial intelligence. *Neural Comput. Appl.* **2021**, *33*, 2497−2514.
(67) Artun, E. Characterizing interwell connectivity in waterflooded reservoirs using data-driven and reduced-physics models: a comparative study. *Neural Comput. Appl.* **2017**, *28*, 1729−1743.
(68) Ghassemzadeh, S.; Gonzalez Perdomo, M.; Haghighi, M.; Abbasnejad, E. A data-driven reservoir simulation for natural gas reservoirs. *Neural Comput. Appl.* **2021**, *33*, 11777−11798.
(69) Al-Gawfi, A.; Zirrahi, M.; Hassanzadeh, H.; Abedi, J. Development of Generalized Correlations for Thermophysical Properties of Light Hydrocarbon Solvents (C1−C5)/Bitumen Systems Using Genetic Programming. *ACS Omega* **2019**, *4*, 6955−6967.
(70) Wang, L.; Tian, Y.; Yao, B.; Yu, X. Machine Learning Analyses of Low Salinity Effect in Sandstone Porous Media. *J. Porous Media* **2020**, *23*, 731−740.
(71) Kondori, J.; Miah, M. I.; Zendehboudi, S.; Khan, F.; Heagle, D. Hybrid connectionist models to assess recovery performance of low salinity water injection. *J. Pet. Sci. Eng.* **2021**, *197*, 107833.
(72) Hidayat, F.; Astsauri, T. M. S. Applied random forest for parameter sensitivity of low salinity water Injection (LSWI) implementation on carbonate reservoir. *Alexandria Eng. J.* **2021**, DOI: 10.1016/j.aej.2021.06.096.
(73) Bernard, G. G. Effect of Floodwater Salinity on Recovery Of Oil from Cores Containing Clays, *SPE California Regional Meeting*, 1967; Los Angeles, California, USA, 1967.
(74) Al-Mumen, A. A.-M. *The Effect of injected water salinity on oil recovery*; King Fahd University of Petroleum and Minerals: Saudi Arabia, 1990.
(75) Yildiz, H. O.; Morrow, N. R. Effect of brine composition on recovery of Moutray crude oil by waterflooding. *J. Pet. Sci. Eng.* **1996**, *14*, 159−168.
(76) Zhou, X.; Morrow, N. R.; Ma, S. Interrelationship of Wettability, Initial Water Saturation, Aging Time, and Oil Recovery by Spontaneous Imbibition and Waterflooding. *SPE J.* **2000**, *5*, 199−207.
(77) McGuire, P. L.; Chatham, J. R.; Paskvan, F. K.; Sommer, D. M.; Carini, F. H. Low Salinity Oil Recovery: An Exciting New EOR Opportunity for Alaska's North Slope, *SPE Western Regional Meeting*, 2005; Irvine, California, USA, 2005.
(78) Zhang, Y.; Morrow, N. R. Comparison of Secondary and Tertiary Recovery With Change in Injection Brine Composition for Crude-Oil/Sandstone Combinations, *SPE/DOE Symposium on Improved Oil Recovery*, Tulsa, Oklahoma, USA, 2006.
(79) Zhang, Y.; Xie, X.; Morrow, N. R. Waterflood Performance By Injection Of Brine With Different Salinity For Reservoir Cores, *SPE Annual Technical Conference and Exhibition*, Anaheim, California, USA, 2007.
(80) Loahardjo, N.; Xie, X.; Yin, P.; Morrow, N. R. Low salinity waterflooding of a reservoir rock, *International Symposium of the Society of Core Analysts*, Calgary, Canada, 2007, pp. 10−12.
(81) Patil, S. B.; Dandekar, A. Y.; Patil, S.; Khataniar, S. Low Salinity Brine Injection for EOR on Alaska North Slope (ANS), *International Petroleum Technology Conference*, Kuala Lumpur, Malaysia, 2008.
(82) Agbalaka, C. C.; Dandekar, A. Y.; Patil, S. L.; Khataniar, S.; Hemsath, J. R. Coreflooding Studies to Evaluate the Impact of Salinity and Wettability on Oil Recovery Efficiency. *Transp. Porous Media* **2009**, *76*, 77.
(83) Soraya, B.; Malick, C.; Philippe, C.; Bertin, H. J.; Hamon, G. Oil Recovery by Low-Salinity Brine Injection: Laboratory Results on Outcrop and Reservoir Cores, *SPE Annual Technical Conference and Exhibition*, New Orleans, Louisiana, USA, 2009.
(84) Alagic, E.; Skauge, A. Combined Low Salinity Brine Injection and Surfactant Flooding in Mixed−Wet Sandstone Cores. *Energy Fuels* **2010**, *24*, 3551−3559.
(85) Ashraf, A.; Hadia, N. J.; Torsæter, O.; Tweheyo, M. T. Laboratory Investigation of Low Salinity Waterflooding as Secondary Recovery Process: Effect of Wettability, *SPE Oil and Gas India Conference and Exhibition*, Mumbai, India, 2010.
(86) Alotaibi, M. B.; Azmy, R. M.; Nasr-El-Din, H. A. A Comprehensive EOR Study Using Low Salinity Water in Sandstone Reservoirs, *SPE Improved Oil Recovery Symposium*, Tulsa, Oklahoma, USA, 2010.
(87) Robertson, E. P. Oil Recovery Increases by Low-Salinity Flooding: Minnelusa and Green River Formations, *SPE Annual Technical Conference and Exhibition*, Florence, Italy, 2010.
(88) Pu, H.; Xie, X.; Yin, P.; Morrow, N. R. Low Salinity Waterflooding and Mineral Dissolution, *SPE Annual Technical Conference and Exhibition*, Florence, Italy, 2010.
(89) Farzaneh, S. A.; Carnegie, A.; Sohrabi, M.; Mills, J. R.; Facanha, J. M.; Sellers, B. A Case Study of Oil Recovery Improvement by Low Salinity Water Injection, *Abu Dhabi International Petroleum Exhibition & Conference*, Abu Dhabi, UAE, 2017.
(90) Cissokho, M.; Bertin, H.; Boussour, S.; Cordier, P.; Hamon, G. Low Salinity Oil Recovery On Clayey Sandstone: Experimental Study. *Petrophys.* **2010**, *51*, 305−313.
(91) Alagic, E.; Spildo, K.; Skauge, A.; Solbakken, J. Effect of crude oil ageing on low salinity and low salinity surfactant flooding. *J. Pet. Sci. Eng.* **2011**, *78*, 220−227.
(92) Hadia, N.; Lehne, H. H.; Kumar, K. G.; Selboe, K.; Stensen, J. Å.; Torsæter, O. Laboratory Investigation of Low Salinity Waterflooding on Reservoir Rock Samples from the Frøy Field, *SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, 2011.
(93) Nasralla, R. A.; Nasr-El-Din, H. A. Impact of Electrical Surface Charges and Cation Exchange on Oil Recovery by Low Salinity Water, *SPE Asia Pacific Oil and Gas Conference and Exhibition*, Jakarta, Indonesia, 2011.
(94) Gamage, P.; Thyne, G. Comparison of Oil Recovery by Low Salinity Waterflooding in Secondary and Tertiary Recovery Modes, *SPE Annual Technical Conference and Exhibition*, Denver, Colorado, USA, 2011.
(95) Nasralla, R. A.; Alotaibi, M. B.; Nasr-El-Din, H. A. Efficiency of Oil Recovery by Low Salinity Water Flooding in Sandstone Reservoirs, *SPE Western North American Region Meeting*, Anchorage, Alaska, USA, 2011.
(96) Thyne, G.; Gamage, P. Evaluation of Low-Salinity Waterflooding for 51 Fields in Wyoming, *Society of Core Analysts, Annual Meeting*, Austin, Texas, USA, 2011.
(97) Hadia, N. J.; Hansen, T.; Tweheyo, M. T.; Torsæter, O. Influence of Crude Oil Components on Recovery by High and Low Salinity Waterflooding. *Energy Fuels* **2012**, *26*, 4328−4335.
(98) Winoto, W.; Loahardjo, N.; Xie, X.; Yin, P.; Morrow, N. R. Secondary and Tertiary Recovery of Crude Oil from Outcrop and Reservoir Rocks by Low Salinity Waterflooding, *SPE Improved Oil Recovery Symposium*, Tulsa, Oklahoma, USA, 2012.
(99) Shaker Shiran, B.; Skauge, A. Wettability and Oil Recovery by Low Salinity Injection, *SPE EOR Conference at Oil and Gas West Asia*, Muscat, Oman, 2012.
(100) Jiazhong, W.; Quan, X.; Li, L.; Qingjie, L.; Jishun, Q.; Desheng, M.; Kangyun, W.; Shouguo, Y., Coreflood study of low salinity effect in low permeability sandstone reservoirs. In *International Symposium of the Society of Core Analysts* Aberdeen, Scotland, UK, 2012.
(101) Romero, M. I.; Gamage, P.; Jiang, H.; Chopping, C.; Thyne, G. Study of low-salinity waterflooding for single- and two-phase experiments in Berea sandstone cores. *J. Pet. Sci. Eng.* **2013**, *110*, 149−154.

(102) Shaker Shiran, B.; Skauge, A. Enhanced Oil Recovery (EOR) by Combined Low Salinity Water/Polymer Flooding. *Energy Fuels* **2013**, *27*, 1223−1235.
(103) Callegaro, C.; Bartosek, M.; Masserano, F.; Nobili, M.; Parracello, V. P.; Pizzinelli, C. S.; Caschili, A. Opportunity of Enhanced Oil Recovery Low Salinity Water Injection: From Experimental Work to Simulation Study up to Field Proposal, *EAGE Annual Conference & Exhibition incorporating SPE Europec*, London, UK, 2013.
(104) Suman, Y. K. *Evaluation of Low Saline "Smart Water" Enhanced Oil Recovery in Light Oil Reservoirs*. University of Regina, Regina, Saskatchewan, 2014.
(105) Hamouda, A. A.; Valderhaug, O. M. Investigating Enhanced Oil Recovery from Sandstone by Low-Salinity Water and Fluid/Rock Interaction. *Energy Fuels* **2014**, *28*, 898−908.
(106) Strand, S.; Austad, T.; Puntervold, T.; Aksulu, H.; Haaland, B.; RezaeiDoust, A. Impact of Plagioclase on the Low Salinity EOR-Effect in Sandstone. *Energy Fuels* **2014**, *28*, 2378−2383.
(107) Shaddel, S.; Tabatabae-Nejad, S. A.; Fathi, S. J. Low-salinity water flooding: Evaluating the effect of salinity on oil and water relative permeability, Wettability, and oil recovery. *Spec. Top, Rev, Reviews in Porous Media: Int. J.* **2014**, *5*, 133−143.
(108) Suijkerbuijk, B. M.; Sorop, T. G.; Parker, A. R.; Masalmeh, S. K.; Chmuzh, I. V.; Karpan, V. M.; Volokitin, Y. E.; Skripkin, A. G. Low Salinity Waterflooding at West Salym: Laboratory Experiments and Field Forecasts, *SPE Improved Oil Recovery Symposium*, Tulsa, Oklahoma, USA, 2014.
(109) Shehata, A. M.; Nasr-El-Din, H. A. Reservoir Connate Water Chemical Composition Variations Effect on Low-Salinity Water-flooding, *Abu Dhabi International Petroleum Exhibition and Conference*, Abu Dhabi, UAE, 2014.
(110) Aghaeifar, Z.; Strand, S.; Austad, T.; Puntervold, T.; Aksulu, H.; Navratil, K.; Storås, S.; Håmsø, D. Influence of Formation Water Salinity/Composition on the Low-Salinity Enhanced Oil Recovery Effect in High-Temperature Sandstone Reservoirs. *Energy Fuels* **2015**, *29*, 4747−4754.
(111) Xie, Q.; Ma, D.; Wu, J.; Liu, Q.; Jia, N.; Luo, M. Potential Evaluation of Ion Tuning Waterflooding for a Tight Oil Reservoir in Jiyuan OilField: Experiments and Reservoir Simulation Results, *SPE Asia Pacific Enhanced Oil Recovery Conference*, Kuala Lumpur, Malaysia, 2015.
(112) Xie, Q.; Ma, D.; Wu, J.; Liu, Q.; Jia, N.; Luo, M. Low Salinity Waterflooding in Low Permeability Sandstone: Coreflood Experiments and Interpretation by Thermodynamics and Simulation, *SPE Asia Pacific Enhanced Oil Recovery Conference*, Kuala Lumpur, Malaysia, 2015.
(113) Callegaro, C.; Bartosek, M.; Nobili, M.; Masserano, F.; Pollero, M.; Baz, D. M.; Kortam, M. M. Design and Implementation of Low Salinity Waterflood in a North African Brown Field, *Abu Dhabi International Petroleum Exhibition and Conference*, Abu Dhabi, UAE, 2015.
(114) Ramanathan, R.; Shehata, A. M.; Nasr-El-Din, H. A. Water Alternating CO2 Injection Process - Does Modifying the Salinity of Injected Brine Improve Oil Recovery?, *OTC Brasil*, Rio de Janeiro, Brazil, 2015.
(115) AlQuraishi, A. A.; AlHussinan, S. N.; AlYami, H. Q. Efficiency and Recovery Mechanisms of Low Salinity Water Flooding in Sandstone and Carbonate Reservoirs, *Offshore Mediterranean Conference and Exhibition*, Ravenna, Italy, 2015.
(116) Piñerez Torrijos, I. D.; Puntervold, T.; Strand, S.; Austad, T.; Abdullah, H. I.; Olsen, K. Experimental Study of the Response Time of the Low-Salinity Enhanced Oil Recovery Effect during Secondary and Tertiary Low-Salinity Waterflooding. *Energy Fuels* **2016**, *30*, 4733−4739.
(117) Hosseinzade Khanamiri, H.; Torsæter, O.; Stensen, J. Å. Effect of Calcium in Pore Scale Oil Trapping by Low-Salinity Water and Surfactant Enhanced Oil Recovery at Strongly Water-Wet Conditions: In Situ Imaging by X-ray Microtomography. *Energy Fuels* **2016**, *30*, 8114−8124.
(118) Liu, Y.; Jiang, T.; Zhou, D.; Zhao, J.; Xie, Q.; Saeedi, A. Evaluation of the Potential of Low Salinity Water Flooding in the High Temperature and High Salinity Dong-He-Tang Reservoir in the Tarim Oilfeild, China: Experimental and Reservoir Simulation Results, *SPE Asia Pacific Oil & Gas Conference and Exhibition*, Perth, Australia, 2016.
(119) Ishiwata, T.; Kurihara, M.; Taniguchi, H.; Tsuchiya, Y.; Watanabe, J.; Matsumoto, K. Investigation on Low Salinity Water-flooding Through Core Flooding Experiments and Numerical Simulation, *SPWLA 22nd Formation Evaluation Symposium of Japan*, Chiba, Japan, 2016.
(120) Mohammad Salehi, M.; Omidvar, P.; Naeimi, F. Salinity of injection water and its impact on oil recovery absolute permeability, residual oil saturation, interfacial tension and capillary pressure. *Egypt. J. Pet.* **2017**, *26*, 301−312.
(121) Alhuraishawy, A. K.; Bai, B.; Wei, M.; Geng, J.; Pu, J. Mineral dissolution and fine migration effect on oil recovery factor by low-salinity water flooding in low-permeability sandstone reservoir. *Fuel* **2018**, *220*, 898−907.
(122) Sohrabi, M.; Mahzari, P.; Farzaneh, S. A.; Mills, J. R.; Tsolis, P.; Ireland, S. Novel Insights Into Mechanisms of Oil Recovery by Use of Low-Salinity-Water Injection. *SPE J* **2017**, *22*, 407−416.
(123) Al-Saedi1, H. N.; Flori, R. E. Experimental Investigation of the Performance of Low Salinity Water Flooding as a Novel Enhanced Oil Recovery. *Sch. J, App, Sci. Res.* **2018**, *1*, 51−55.
(124) Suleimanov, B. A.; Latifov, Y. A.; Veliyev, E. F.; Frampton, H. Comparative analysis of the EOR mechanisms by using low salinity and low hardness alkaline water. *J. Pet. Sci. Eng.* **2018**, *162*, 35−43.
(125) Al-Saedi, H. N.; Flori, R. E. Enhanced oil recovery of low salinity water flooding in sandstone and the role of clay. *Pet. Explor. Dev.* **2018**, *45*, 927−931.
(126) Zhang, L.; Zhang, J.; Wang, Y.; Yang, R.; Zhang, Y.; Gu, J.; Zhang, M.; Ren, S. Experimental Investigation of Low-Salinity Water Flooding in a Low-Permeability Oil Reservoir. *Energy Fuels* **2018**, *32*, 3108−3118.
(127) Abdelmoneim, S. S.; Nasr-El-Din, H. A. Effect of Low Salinity Waterflooding on Sweep Efficiency in Multilayered Clay-Rich Sandstone Reservoirs, *SPE Improved Oil Recovery Conference*, Tulsa, Oklahoma, USA,, 2018.
(128) Rock, A.; Hincapie, R. E.; Hoffmann, E.; Ganzer, L. Tertiary Low Salinity Waterflooding LSWF in Sandstone Reservoirs: Mechanisms, Synergies and Potentials in EOR Applications, *SPE Europec featured at 80th EAGE Conference and Exhibition*, Copenhagen, Denmark, 2018.
(129) Almeida, A.; Patel, R.; Arambula, C.; Trivedi, J.; Soares, J.; Costa, G.; Embiruçu, M. Low Salinity Water Injection in a Clastic Reservoir in Northeast Brazil: An Experimental Case Study, *SPE Trinidad and Tobago Section Energy Resources Conference*, Port of Spain, Trinidad and Tobago, 2018.
(130) Aghaeifar, Z.; Puntervold, T.; Strand, S.; Austad, T.; Maghsoudi, B.; Ferreira, J. D. Low Salinity EOR Effects After Seawater Flooding in a High Temperature and High Salinity Offshore Sandstone Reservoir, *SPE Norway One Day Seminar*, Bergen, Norway, 2018.
(131) Al-Saedi, H. N.; Flori, R. E.; Mortadha, A., Coupling Low Salinity Water Flooding and Steam Flooding for Heavy Oil in Sandstone Reservoirs; Low Salinity-Alternating-Steam Flooding LSASF: A Novel EOR Technique. In *Abu Dhabi International Petroleum Exhibition & Conference*, Abu Dhabi, UAE, 2018.
(132) Al-Saedi, H.; Williams, A.; Di Torino, P.; Flori, R. Comparison Between Cold/Hot Smart Water Flooding in Sandstone Reservoirs. *J. Pet. Environ. Biotechnol.* **2018**, *9*, 1000365.
(133) Al-Saedi, H. N.; Flori, R. E.; Brady, P. V. Insights into Low Salinity Water Flooding, *Oral presentation given at the International Symposium of the Society of Core Analysts*, Society of Petrophysicists and Well-Log Analysts: Trondheim, Norway, 2018; pp. 27−30.
(134) Al-Saedi, H. N.; Flori, R. E.; Alkhamis, M.; Brady, P. V. Coupling of Low-Salinity Water Flooding and Steam Flooding for

https://doi.org/10.1021/acsomega.1c05493
ACS Omega 2021, 6, 32304−32326

Sandstone Unconventional Oil Reservoirs. *Nat. Resour. Res.* **2019**, *28*, 213−221.
(135) Khishvand, M.; Oraki Kohshour, I.; Alizadeh, A. H.; Piri, M.; Prasad, S. A multi-scale experimental study of crude oil-brine-rock interactions and wettability alteration during low-salinity water-flooding. *Fuel* **2019**, *250*, 117−131.
(136) Al-Saedi, H. N.; Flori, R. E.; Brady, P. V. Effect of divalent cations in formation water on wettability alteration during low salinity water flooding in sandstone reservoirs: Oil recovery analyses, surface reactivity tests, contact angle, and spontaneous imbibition experiments. *J. Mol. Liq.* **2019**, *275*, 163−172.
(137) Al-Saedi, H. N.; Brady, P. V.; Flori, R. E.; Heidari, P. Insights into the role of clays in low salinity water flooding in sand columns. *J. Pet. Sci. Eng.* **2019**, *174*, 291−305.
(138) Mokhtari, R.; Ayatollahi, S.; Fatemi, M. Experimental investigation of the influence of fluid-fluid interactions on oil recovery during low salinity water flooding. *J. Pet. Sci. Eng.* **2019**, *182*, 106194.
(139) Ding, Y.; Zheng, S.; Meng, X.; Yang, D. Low Salinity Hot Water Injection With Addition of Nanoparticles for Enhancing Heavy Oil Recovery. *J. Energy Resour. Technol.* **2019**, *141*, No. 072904.
(140) Al-Saedi, H. N.; Flori, R. E. Novel Coupling Smart Water-CO2 Flooding for Sandstone Reservoirs. *Petrophys.* **2019**, *60*, 525−535.
(141) Almeida da Costa, A.; Trivedi, J.; Soares, J.; Rocha, P.; Costa, G.; Embiruçu, M. An experimental evaluation of low salinity water mechanisms in a typical Brazilian sandstone and light crude oil with low acid/basic number. *Fuel* **2020**, *273*, 117694.
(142) Sagala, F.; Hethnawi, A.; Nassar, N. N. Integrating Silicate-Based Nanoparticles with Low-Salinity Water Flooding for Enhanced Oil Recovery in Sandstone Reservoirs. *Ind. Eng. Chem. Res.* **2020**, *59*, 16225−16239.
(143) Kakati, A.; Kumar, G.; Sangwai, J. S. Oil Recovery Efficiency and Mechanism of Low Salinity-Enhanced Oil Recovery for Light Crude Oil with a Low Acid Number. *ACS Omega* **2020**, *5*, 1506−1518.
(144) Zhao, Y.; Yin, S.; Seright, R. S.; Ning, S.; Zhang, Y.; Bai, B. Performance of Low Salinity Polymer Flood in Enhancing Heavy Oil Recovery on the Alaska North Slope, *Proceedings of the 8th Unconventional Resources Technology Conference*, Virtual, 2020; Unconventional Resources Technology Conference (URTeC), pp 1929−1943.
(145) Al-Saedi, H. N.; Al-Jaberi, S. K.; Flori, R. E.; Al-Bazzaz, W. Novel Insights into Low Salinity Water Flooding Enhanced Oil Recovery in Sandstone: The Role of Calcite, *SPE Improved Oil Recovery Conference*, Virtual, 2020.
(146) Zhao, Y.; Yin, S.; Seright, R. S.; Ning, S.; Zhang, Y.; Bai, B. Enhancing Heavy-Oil-Recovery Efficiency by Combining Low-Salinity-Water and Polymer Flooding. *SPE J.* **2021**, *26*, 1535−1551.
(147) Maya-Toro, G.-A.; Cardona, L.; Rueda, M.; Cortés, F. Effect of ionic strength in low salinity water injection processes. *C.T.F Cienc. Tecnol. Futuro* **2020**, *10*, 17−26.
(148) Tahir, M.; Hincapie, R. E.; Ganzer, L. Influence of Sulfate Ions on the Combined Application of Modified Water and Polymer Flooding—Rheology and Oil Recovery. *Energies* **2020**, *13*, 2356.
(149) Bhicajee, P.; Romero-Zerón, L. Effect of different low salinity flooding schemes and the addition of alkali on the performance of low-salinity waterflooding during the recovery of heavy oil from unconsolidated sandstone. *Fuel* **2021**, *289*, 119981.
(150) Snosy, M. F.; El Ela, M. A.; El-Banbi, A.; Sayyouh, H. Impact of the injected water salinity on oil recovery from sandstone formations: Application in an Egyptian oil reservoir. *Petroleum* **2021**, DOI: 10.1016/j.petlm.2021.03.006.
(151) Almeida da Costa, A.; Mateo, J.; Patel, R.; Trivedi, J. J.; Soares, J. B. P.; Rocha, P. S.; Costa, G.; Embiruçu, M. Can low salinity water injection be an efficient option for the Recôncavo Basin? An experimental and simulation case study. *J. Pet. Sci. Eng.* **2021**, *202*, 108557.
(152) Mahboubi Fouladi, M.; Hassani, K.; Rostami, B.; Pourafshary, P. Experimental Studies of Low Salinity Water Flooding in Sandstone Porous Media: Effects of the Presence of Silica and Kaolinite. *Energy Sources, Part A* **2021**, 1−17.
(153) Bilogur, A. Missingno: a missing data visualization suite. *J. Open Source Software* **2018**, *3*, 547.
(154) Bergstra, J.; Komer, B.; Eliasmith, C.; Yamins, D.; Cox, D. D. Hyperopt: a Python library for model selection and hyperparameter optimization. *Comput. Sci. Discovery* **2015**, *8*, No. 014008.
(155) Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694−701.
(156) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. learn. Res.* **2011**, *12*, 2825−2830.
(157) Hornik, K.; Stinchcombe, M.; White, H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Networks* **1990**, *3*, 551−560.
(158) Hemmati-Sarapardeh, A.; Ameli, F.; Dabir, B.; Ahmadi, M.; Mohammadi, A. H. On the evaluation of asphaltene precipitation titration data: Modeling and data assessment. *Fluid Phase Equilib.* **2016**, *415*, 88−100.
(159) Suykens, J. A.; Van Gestel, T.; De Brabanter, J. *Least squares support vector machines*. World scientific: 2002, DOI: 10.1142/5089.
(160) Nilsson, N. J., *Learning machines: foundations of trainable pattern-classifying systems*. 1st Edition ed.; McGraw-Hill Companies: 1965.
(161) Tatar, A.; Shokrollahi, A.; Halali, M. A.; Azari, V.; Safari, H. A hybrid intelligent computational scheme for determination of refractive index of crude oil using SARA fraction analysis. *Can. J. Chem. Eng.* **2015**, *93*, 1547−1555.
(162) Haykin, S., *Neural Networks: A Comprehensive Foundation*. Pearson Education, Inc.: 1999; p 823.
(163) Perrone, M. P.; Cooper, L. N. *When networks disagree: Ensemble methods for hybrid neural networks*; 1992; p 15.
(164) Rousseeuw, P. J.; Leroy, A. M., *Robust regression and outlier detection*. John wiley & sons: 2005; Vol. *589*.
(165) scikit-learn 1.10. Decision Trees.
(166) Hastie, T.; Tibshirani, R.; Friedman, J., *The elements of statistical learning: data mining, inference, and prediction. Springer series in statistics*. 2009, DOI: 10.1007/978-0-387-84858-7.
(167) Jadhunandan, P. P. *Effects of brine composition, crude oil, and aging conditions on wettability and oil recovery*. Department of Petroleum Engineering, New Mexico Institute of Mining & Technology, 1990.
(168) Qi, Z.; Wang, Y.; He, H.; Li, D.; Xu, X. Wettability Alteration of the Quartz Surface in the Presence of Metal Cations. *Energy Fuels* **2013**, *27*, 7354−7359.
(169) Al-Nofli, K.; Pourafshary, P.; Mosavat, N.; Shafiei, A. Effect of initial wettability on performance of smart water flooding in carbonate reservoirs—an experimental investigation with ior implications. *Energies* **2018**, *11*, 1394.