ACS OMEGA

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

Article

Asphaltene Precipitation Prediction during Bitumen Recovery: Experimental Approach versus Population Balance and Connectionist Models

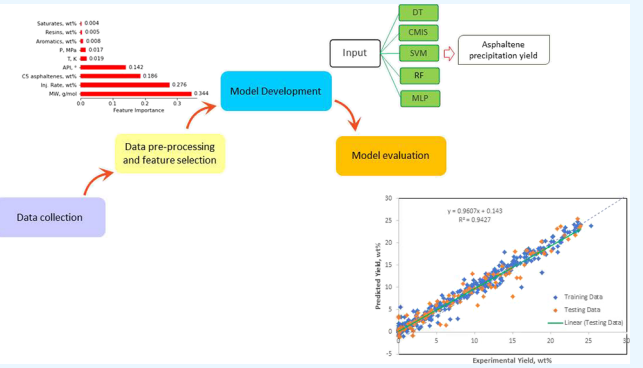
Turar Yerkenov, Simin Tazikeh, Afshin Tatar, and Ali Shafiei*

Cite This: *ACS Omega* 2022, 7, 33123–33137

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Deasphalting bitumen using paraffinic solvent injection is a commonly used technique to reduce both its viscosity and density and ease its flow through pipelines. Common modeling approaches for asphaltene precipitation prediction such as population balance model (PBM) contains complex mathematical relation and require conducting precise experiments to define initial and boundary conditions. Machine learning (ML) approach is considered as a robust, fast, and reliable alternative modeling approach. The main objective of this research work was to model the effect of paraffinic solvent injection on the amount of asphaltene precipitation using ML and PBM approaches. Five hundred and ninety (590) experimental data were collected from the literature for model development. The gathered data was processed using box plot, data scaling, and data splitting. Data pre-processing led to the use of 517 data points for modeling. Then, multilayer perceptron, random forest, decision tree, support vector machine, committee machine intelligent system optimized by annealing, and random search techniques were used for modeling. Precipitant molecular weight, injection rate, API gravity, pressure, C_5 asphaltene content, and temperature were determined as the most relevant features for the process. Although the results of the PBM model are precise, the AI/ML model (CMIS) is the preferred model due to its robustness, reliability, and relative accuracy. The committee machine intelligent system is the superior model among the developed smart models with an RMSE of 1.7% for the testing dataset and prediction of asphaltene precipitation during bitumen recovery.



1. INTRODUCTION

Petroleum makes up a notable portion of the global energy basket to deal with growing energy demands and economic development. At the same time, sources of conventional crude oil (e.g., light crude oils) are limited and are on decline. Hence, the petroleum industry has attempted to recover unconventional crude oil such as heavy oil, extra heavy oil, bitumen, oil shale, and shale oil to address the demand.[1] For instance, Canada significantly increased its bitumen production compared to light crude oil in the past five decades. Various daily crude oil production types in Canada from 1971 to 2019 are shown in Figure 1.

The in situ viscosity of bitumen is very high (10,000 to 500,000 mPa·s at 25 °C) and raises many issues associated with production and pipeline transportation. Injection of naphtha and paraffinic solvents (anti-solvent) are two main commercial methods for decreasing bitumen viscosity.[3] Naphtha dilutes bitumen and decreases its density and viscosity. Paraffinic solvents precipitate the heavy bitumen component (e.g., asphaltene) and help bitumen to flow easily. The yield of the paraffinic solvent method in bitumen recovery is low compared to the naphtha injection method. Nevertheless, the chance of coke formation in the paraffinic solvent method is very low because of the rejection of asphaltene from bitumen. In addition, the asphaltene precipitation phenomenon was observed during conventional crude oil during oil recovery, production, transportation, and storage when equilibrium conditions (e.g., temperature, pressure, and composition) are disturbed.

Asphaltene is the most polar and the heaviest component of crude oils, having a sophisticated molecular architecture that is defined in terms of solubility in different substances.[4−7] Asphaltenes can be dissolved in aromatic hydrocarbons (e.g., toluene and benzene), and it is not soluble in paraffinic solvent with low molecular weight, including *n*-hexane, *n*-heptane, and *n*-pentane.[4,5,8] Asphaltene aggregation increases the crude oil

Received: May 25, 2022
Accepted: August 24, 2022
Published: September 9, 2022

© 2022 The Authors. Published by American Chemical Society

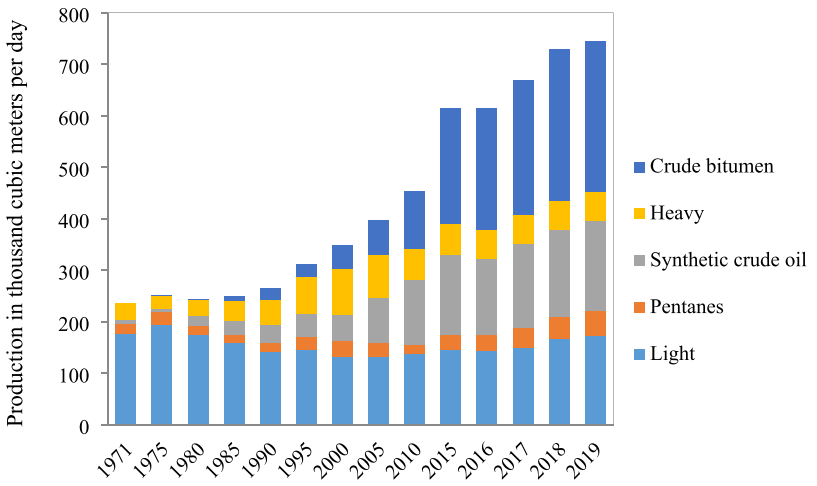


Figure 1. Daily crude oil production types in Canada from 1971 to 2019. Data courtesy of statista.com.[2]

viscosity and promotes the stability of water-in-oil emulsions,[8] implying the need for removing them from crude oil. Asphaltene precipitation can result in undesired wettability alteration, plugging wellbores,[9] and permeability reduction[10] leading to flow assurance issues,[7] equipment damage, increased capital/operation costs, and decreased production.[11,12]

Asphaltene precipitation usually occurs during some EOR methods such as miscible gas injection and solvent injection because of compositional modifications and pressure drop.[5,7,10,12] The nature of solvent/precipitant injected into crude oil directly affects the amount of asphaltene precipitation. For instance, the naphthenic and resin groups in the solvent structure have the same effect as aromatic compounds and inhibit asphaltene precipitation. In contrast, adding paraffinic solvents (e.g., *n*-butane, *n*-hexane, and *n*-heptane) leads to asphaltene precipitation.[6] Hence, the nature of solvent and its composition for asphaltene deposition prevention is very important. For establishing this aim, Khormali et al.[13] introduced a new chemical solvent package for removing asphaltene from the system. Toluene, pyridine, methanol, surfactant dodecyl benzene sulfonic, and sodium hydroxide are the main elements in the new solvent. Their results showed that the permeability of the core improved from 71 to 94% after the use of new solvent in comparison with toluene. In another research work, the composition of polyacrylamide, amphoteric fatty acid, ethylene copolymer, maleic anhydride, organic sulfate, and 1-hydroxyethane-1,1-diphosphonic acid was used as an asphaltene inhibitor by Khormali et al.[9] The IFT between crude oil and inhibitors decreases with increasing organic sulfate, and 6% of organic sulfate was found as the optimum concentration for decreasing the IFT between crude oil and inhibitors, meanwhile, the concentration of other components was held constant. The isolation of asphaltenes out of the crude oil mixture is a significant procedure for decreasing bitumen viscosity, measuring characteristics of asphaltenes, and examining aggregation and deposition mechanisms. Hence, estimation of the asphaltene precipitation amount because of the addition of a precipitant is essential before planning the production operations.[14] The conventional method to isolate asphaltene from crude oil is the addition of a precipitant (anti-solvent) to the crude oil sample.[7,15] This method enables investigation of the asphaltene precipitation amount. However, the asphaltene aggregation mechanism to understand the precipitation phenomenon is not well understood, is still being debated, and needs to be further clarified.

Different research works are reported in the literature to model asphaltene aggregation process from different viewpoints.[16−18] In general, thermodynamic and kinetic approaches are two main methods to model asphaltene aggregation. Thermodynamic models categorize into two main groups: solubility and colloidal approaches.[19] These methods are based on the asphaltene state in crude oil. The solubility model assumes that crude oil is divided into two phases: asphaltene and de-asphaltene phases, while asphaltene surrounded by resin is considered in the colloidal model.[20] These models can help to enhance the knowledge of asphaltene aggregation/precipitation phenomena. The population balance model (PBM) is a kinetic model to explain the asphaltene aggregation mechanism.[21] This model helps to predict the size distribution of asphaltene aggregates under unstable thermodynamic conditions. Nucleation and growth of asphaltene aggregates are two main steps considered in the PBM. In addition, generation of new asphaltene aggregates (birth process) and breakage ones (death process) during the time are considered in the PBM equations that help to predict the asphaltene kinetic properly.[22] The common PBM can predict size distribution of the asphaltene particles but cannot calculate the amount of asphaltene precipitation. Hence, it is vital to use mass balance models at the same time to predict the amount of asphaltene precipitation. Khoshandam and Alamdari[21] used the PBM and mass equations to model asphaltene aggregation inside a heptane−toluene solution. They considered the asphaltene growth and agglomeration mechanisms as an exponential dependency to supersaturation. In addition, the kinetic parameters of the model were identified using experimental data to match the modeling results. In their model, there was 10% of relative deviation between modeling results and experimental data. Moradi et al.[22] investigated the size distribution of asphaltene aggregates during miscible gas injection and natural depletion using the PBM. They concluded that the PBM can be used to determine the size distribution of asphaltene particles effectively, and it accounts for the aggregation mechanism to optimize the collision factor. They reported that one optimal collision factor is necessary for unimodal distribution and two different optimal collision factors are required for bimodal curves. Duran[1] applied the PBM to the kinetics of asphaltene precipitation upon adding *n*-heptane for

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

four types of heavy crude oil. They considered diameters of asphaltene aggregates and precipitation yield measured at anaerobic and room conditions. In addition, they used the modified model to enhance the accuracy of their work. Nassar et al.[23] used a PBM to show the effect of nanoparticles (e.g., magnetite, alumina, and commercial silica) in preventing asphaltene precipitation. They reported that the hydrodynamic size of asphaltene aggregates decreases in the presence of nanoparticles. Their prediction by the PBM was in line with experimental results with acceptable error. Besides the accuracy of the PBM in predicting the amount of asphaltene precipitation, this method contains complex mathematical relation and requires conducting precise experiments to define the initial and boundary conditions. Performing experiments requires significant time, specific equipment, and financial resources. In addition, providing reservoir thermodynamic conditions in the lab is challenging and complicates conducting experiments under high-pressure, high-temperature (HPHT) conditions. Moreover, solving complex equations and knowing the nature of the reactions are two main limitations in theoretical methods. Therefore, development of robust, accurate, fast, and reliable predictive tools are inevitable.

Connectionist or smart computational tools can provide a practical way to overcome these limitations. Artificial intelligence/machine learning (AI/ML) approaches can connect input parameters and target values using a training process. They can also estimate the parameters that were not considered in the network.[24] Artificial neural network (ANN), genetic algorithm (GA), support vector machine (SVM) and support vector regression (SVR), least-squares support vector machine (LSSVM), decision tree (DT), random forest (RF), and particle swarm optimization (PSO) are the most common AI models for estimating the amount of asphaltene precipitation that are used in the previous research works. For instance, Sattari et al.[25] developed AI models such as decision tree (DT), gene expression programming (GEP), and least squares support vector machine (LSSVM) to estimate the amount of asphaltene precipitation from an Iranian crude oil. The independent parameters for their research work included the temperature, *n*-alkane molecular weight, and *n*-alkane over oil ratio. GEP (AARD = 8.5%) and LSSVM (AARD = 3%) methods (in contrast to DT, AARD = 11%) demonstrated excellent performance strongly correlated to the obtained experimental data. In another research work, Bassir and Madani[26] used the LSSVM method to predict the asphaltene precipitation weight percentage in the presence of three types of paraffinic solvents such as C_{22}, *n*-C_{10}, and commercial paraffinic pool. The model was optimized by a coupled simulated annealing (CSA) technique. They used a 111 experimental data sample database to develop their AI model. Their modeling results with AARD 0.068% showed a good agreement with the experimental data obtained from titration tests. Ghorbani et al.[27] developed a support vector regression (SVR) model to forecast the amount of precipitated asphaltene. They used generic algorithm (GA) to optimize the performance of the SVM model. Their suggested model performed better than scaling models for various datasets from the literature,[28,29] as was evident from graphical and statistical accuracy assessments. Based on these valuable results, we can calculate the amount of asphaltene precipitation during bitumen recovery using AI/ML techniques.

To the best of our knowledge, this is the first application of AI/ML models to develop connectionist models for asphaltene precipitation deposition during bitumen recovery by injection of paraffinic solvents. The outcomes of the present work could validate other experimental research and help to further optimize the process and reduce operation/production costs. In this paper, various connectionist models such as DT, RF, SVM, CMIS, and MLP were developed to predict asphaltene precipitation due to paraffinic solvent injection during bitumen recovery. The structure of this paper is as follows. First, data gathering and data pre-processing techniques such as removal of duplicates and low-variance features, missing value imputation, collinearity assessment, data characteristics, outlier removal, feature selection, data scaling, and data splitting are explained. Then, the model development stage and producer are described. After that, the outcomes of models are presented and discussed in detail. Finally, the main conclusions, practical implications, limitations, and some suggestions for future research are presented.

2. DATA COLLECTION AND PRE-PROCESSING

2.1. Data Collection. The first stage of the research involved developing a comprehensive and accurate database. The literature on this subject was reviewed, and parameters were determined based on the physics of the process. Overall, 590 data samples were gathered from eight different studies.[1,6,14,30−34] The input parameters were categorized into several groups, such as operational parameters (five features), oil parameters (eight features), and anti-solvent properties (one feature). The asphaltene precipitation yield (%) was selected as a target parameter. The pressure (MPa), temperature (K), anti-solvent/oil (v/v), anti-solvent/oil (cm^3/g), and anti-solvent rate injection (%) were defined as operational parameters. Oil parameters included API gravity (degrees), density (kg/m^3), viscosity (cP), SARA content (%), and water content (%). Molecular weight (g/mol) was recognized as the primary characteristic of anti-solvent. The effect of each parameter on the asphaltene aggregation process and consequently asphaltene precipitation is different. For instance, the chance of asphaltene aggregation before bubble point pressure will increase with the increase in pressure; meanwhile, there is a reverse trend after the bubble point.

2.2. Data Pre-Processing. *2.2.1. Removal of Duplicates and Low-Variance Features.* Duplicates in the dataset create an unnecessary load for processing. Two types of duplicates can be recognized in the dataset by monitoring input and output variables. Type I duplicates implies the same values for both input and output variables. Whereas, type II duplicates mean that different output variables are retrieved for the identical input variables. For this study, 77 duplicates were eliminated manually based on the authors' judgment before other pre-processing procedures. Apart from the removal of duplicates, the removal of low-variance features is regarded as a necessary stage of data cleaning or quality control. These are defined as features for which most values are identical. Since they have a trivial impact on the dependent parameter, they need to be eliminated. As the variance is defined as a distance-based quantity, the data must be normalized between 0 and 1. A threshold is necessary to define when a feature becomes a low-variance one. In this research work, the threshold was set at variance values less than 0.5% (0.005) of the number of data samples. Zero features were recognized as low-variance features.

2.2.2. Missing Value Imputation. It is not always feasible to assemble a database from experiments for which data for all features are available. There are two main approaches to this issue: dropping the value or handling it using imputation. The

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

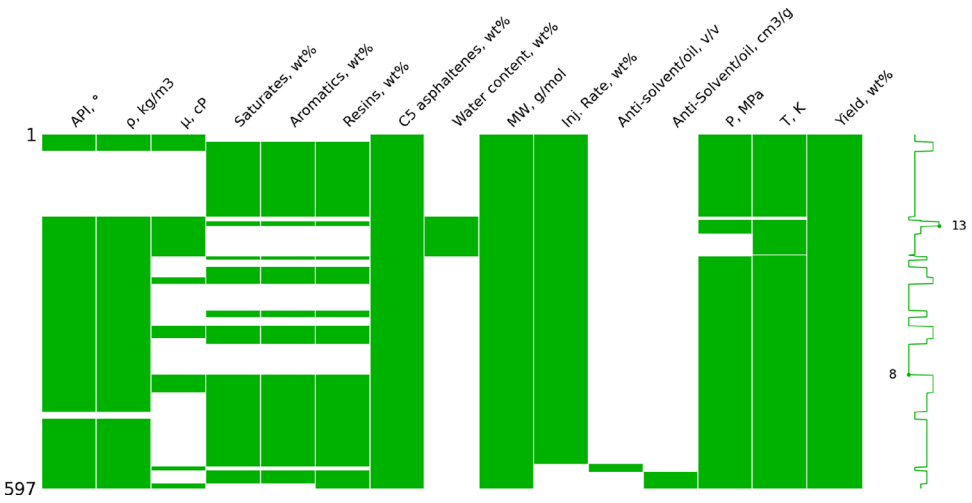


Figure 2. Illustration of absent values in the database.

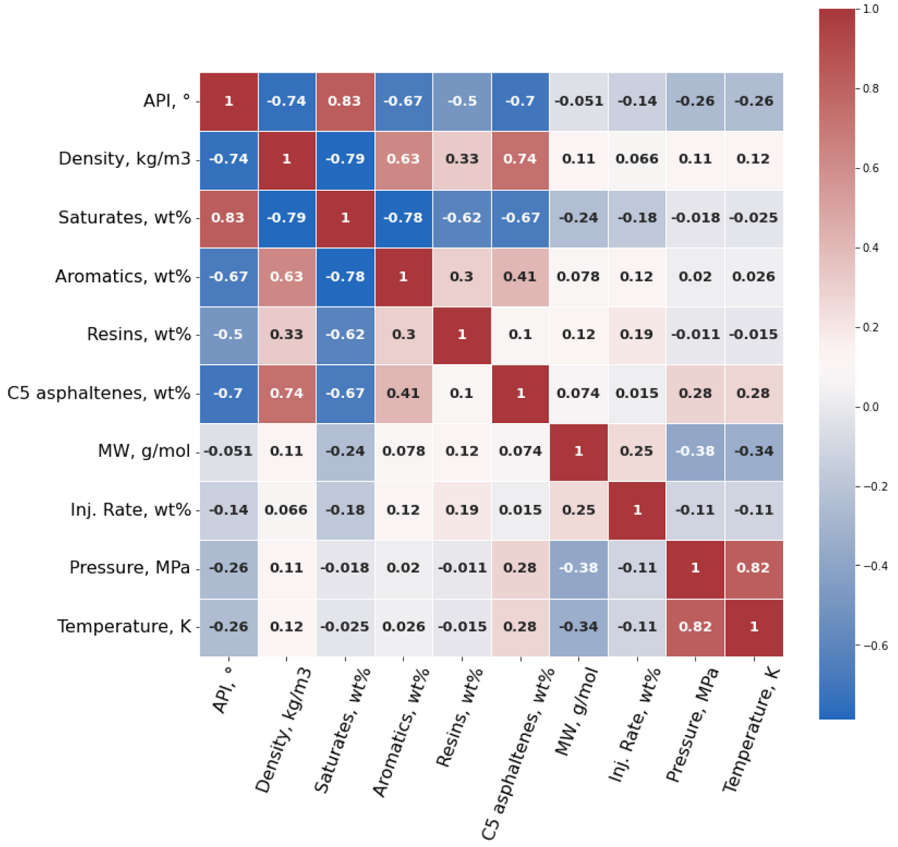


Figure 3. Collinearity assessment using the Pearson correlation coefficient matrix heat map.

loss of knowledge resulting from dropping of the values leads to inefficient learning for the models, subsequently. Hence, filling up these missing values is of paramount importance. Missing value imputation can be performed using a middle value (mean, mode, or median). The mean was used as an imputation tool for the current research work. Prior to imputation, it is necessary to ensure that there is a sufficient amount of data for imputation. If a column has more than 50% of data missing, then it is dropped, and the attempt for imputation is dismissed.[35,36] Missingno and Bilogur's[37] method was used to illustrate the missing values. Missing values from the database are shown in Figure 2. Viscosity, water content, anti-solvent/oil (v/v), and anti-solvent/oil (cm^3/g) were omitted. Imputation was conducted on variables such as the API degree, density, saturates, aromatics, resins, pressure, and temperature. After that, a low-variance feature removal was performed and 4 type I and 65 type II duplicates were subsequently eliminated.

2.2.3. Collinearity Assessment. Collinear input variables do not to facilitate modeling efficiently. Their presence causes repetitive data leading to the increase in size of the database unnecessarily, which results in additional inessential computational load. Since for all features the data distribution was not normal, a heat map was constructed using the Pearson correlation coefficient (R) to assess potential mutual collinearity

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

between two features. The range of the coefficient is between −1 and 1.[38] The features are considered collinear when the absolute value of R is above the threshold determined as 0.9 for this research work. The next step involves identifying features that should remain and the ones that should be dropped using the variance inflation factor (VIF). This factor assesses the multicollinearity of the input parameter; therefore, the one with a higher value needs to be discarded. The heat map matrix constructed using the Pearson correlation coefficient is shown in Figure 3. It is apparent that the API gravity and density were collinear. Therefore, the VIF needs to be observed to determine which one to exclude. Density and API gravity VIF magnitudes were equal to 541.86 and 13.46, respectively. As a result, density was selected to be removed from the database. As a result, after elimination of the density, the VIF values for all variables experienced an evident reduction with an average VIF ranging from 104.77 to 41.48 as indicated in Table 1.

Table 1. VIF Values for Input Parameters Before and After Discarding Collinear Features

Features	VIF before	VIF after
API, °	13.46	10.69
ρ, kg/m^3	541.86	-
saturates, wt %	52.63	26.02
aromatics, wt %	124.96	73.18
resins, wt %	50.49	26.35
C_5 asphaltenes, wt %	38.67	20.93
MW, g/mol	17.63	17.55
inj. rate, wt %	27.86	27.53
P, MPa	5.43	5.35
T, K	174.71	165.74

2.2.4. Data Characteristics. Various statistical parameters describing data in this research work are provided in Table 2. The RMSE was used to show the error of prediction. The prediction error for different numbers of features is shown in Figure 4. The green rectangle, green horizontal line, and red asterisk show average, median, and fliers, respectively. The length of the blue box shows the interquartile range (IQR). The values used in the box plot were standardized to prevent the influence of diversity of raw data using the equation below:

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

where $X^i_{\text{Standard}}$is the standardized value, $X^i_{\text{Original}}$is the original value, $\overline{X}_{\text{Original}}$is the average of the original values, and $\text{SD}_{\text{Original}}$ is the standard deviation for the original values.

2.2.5. Outlier Removal. Outliers recognized during pre-processing stage are called interesting outliers. They are data points remote from the most, yet they are not inherently erroneous. Their removal is favorable as they ruin the normal distribution of data. If these points are included in the testing dataset, they will yield inadequate modeling results. Five times the standard deviation (SD) was used as a threshold to identify outliers in this research work. Consequently, four interesting outliers were dismissed from the database.

2.2.6. Feature Selection. Features used for modeling must be relevant to avoid overfitting problems. The feature selection process was used using recursive feature elimination (RFE). Feature relevance can be determined using the decision tree (DT) process. In this procedure, the feature having minimal relevance needs to be dismissed step-by-step until the germane number of features is left. Consequently, the number of features was reduced from nine to six, with three irrelevant features removed as indicated in Figure 5. RMSE values for features are illustrated as a box plot in Figure 6. The most relevant features include pressure, temperature, API gravity, C_5 asphaltene content, injection rate, and molecular weight of precipitants. These features obtained are consistent with those from experimental results as they can significantly affect the asphaltene precipitation yield in contrast to other parameters. Hence, they should be gathered and used to identify the yield for AI/ML modeling.

2.2.7. Data Splitting. The data was divided into a 4:1 ratio for training (80%) and testing (20%) phases. The training dataset is used to identify parameters and hyperparameters of modeling. At the same time, the testing dataset is used to assess the modeling performance. The overfitting is regulated using 10-fold cross-validation.

2.2.8. Data Scaling. The different scales for the modeling features can cause a disproportionate impact from them. This can be fixed by applying the data scaling. This leads to data to have values of 0 and 1 for the mean and standard deviation, respectively.

3. MODEL DEVELOPMENT AND COMPUTATIONAL PROCEDURE

In this section, the computational procedures conducted in this research work are presented. In addition, the PBM is introduced to evaluate the accuracy of the AI/ML models developed in this research work. The flowchart of the modeling steps taken is presented in Figure 7.

3.1. Multilayer Perceptron. Multilayer perceptron (MLP) is one of the most prevalent feedforward ANNs (artificial neural networks) owing to its ease of execution. This neural network is comprised of input, output, and hidden layers with a certain number of neurons in each.[39] The hidden layer receives the information from the input layer and delivers it to the output layer. The number of neurons in input and output layers is specified by the number of input and output variables, respectively. The number of neurons in hidden layers is random

Table 2. Statistical Parameters for All Features in This Research Work

Parameter	Min	Q_1	Median	Q_3	Max	Mean	Mode	S	k	Range	IQR	SD
anti-solvent MW, g/mol	44.1	72.15	100.2	100.2	194.64	90.86	100.2	1.56	5.39	150.54	100.2	23.29
inj. rate, wt %	31.36	59.42	70	80.08	97.52	69.81	80	−0.11	−0.74	66.16	70	14.18
C_5 asphaltenes, wt %	3.5	14.6	19.4	20.4	24.2	16.52	20.1	−1.18	0.13	20.7	19.4	6.06
API, °	0.01	6.81	8.35	10.85	33.89	10.18	10.85	1.74	2.88	33.88	8.35	8.45
T, K	273	294	294	298	523	312.16	294	2.63	6.75	250	294	43.79
P, MPa	0.1	0.1	0.1	0.1	13.8	1.41	0.1	2.35	4.4	13.7	0.1	2.98
asphaltene yield, wt %	0	2.08	6.8	13.47	25.35	8.35	0	0.56	−0.8	25.35	7.06	6.95

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

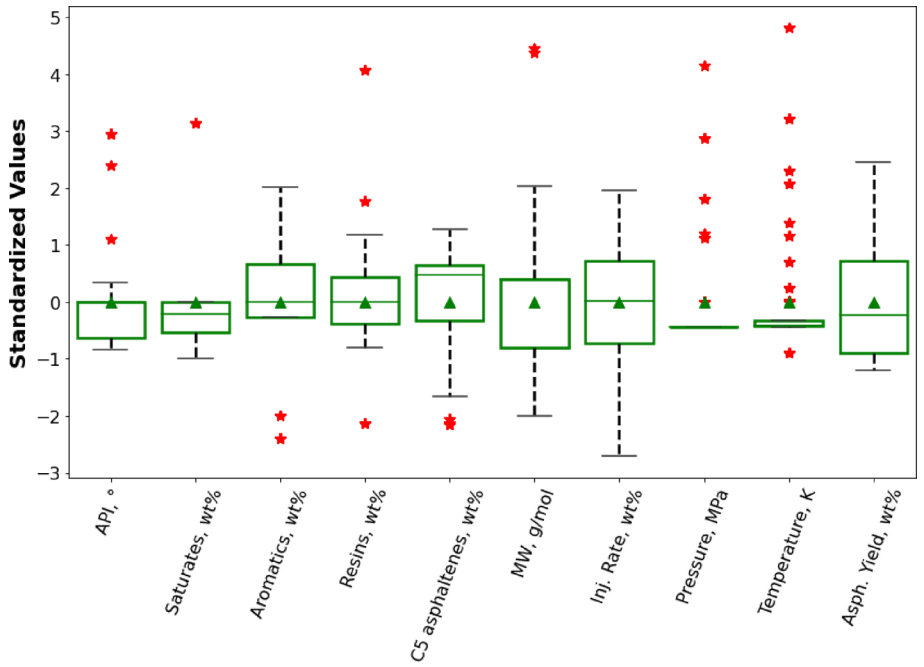


Figure 4. Standardized values for data in the present research work.

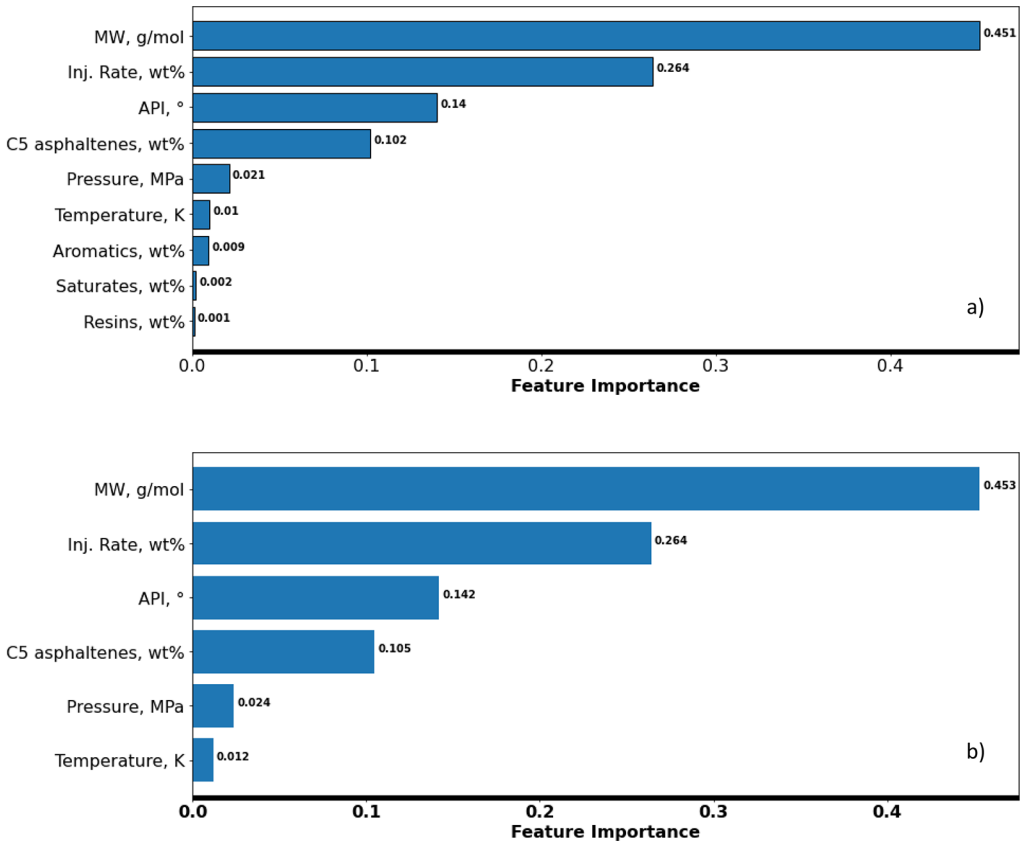


Figure 5. Feature importance of the decision tree model developed using (a) all and (b) the six independent features.

and is usually optimized by trial and error. The model is developed until the difference between desired and actual values is minimized.

The scikit-learn library offers numerous hyperparameters necessary for execution of the MLP. If the values are not adjusted manually by the user, the default values will automatically be used. The number of hidden layers and the number of neurons in them are considered essential hyperparameters. The solver is presented on "lbfgs". As shown in Figure 8 and Table 3, the RS and AL optimization methods converge to the same values, thereby exhibiting identical performance. The best error for the best MLP architecture after 100 iterations is equal to 0.261.

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

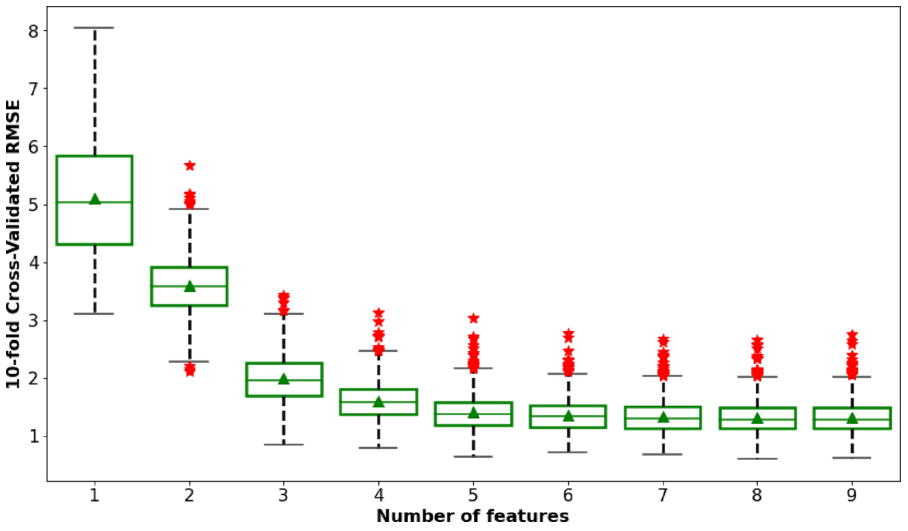


Figure 6. RMSE box plot developed by various numbers of features.

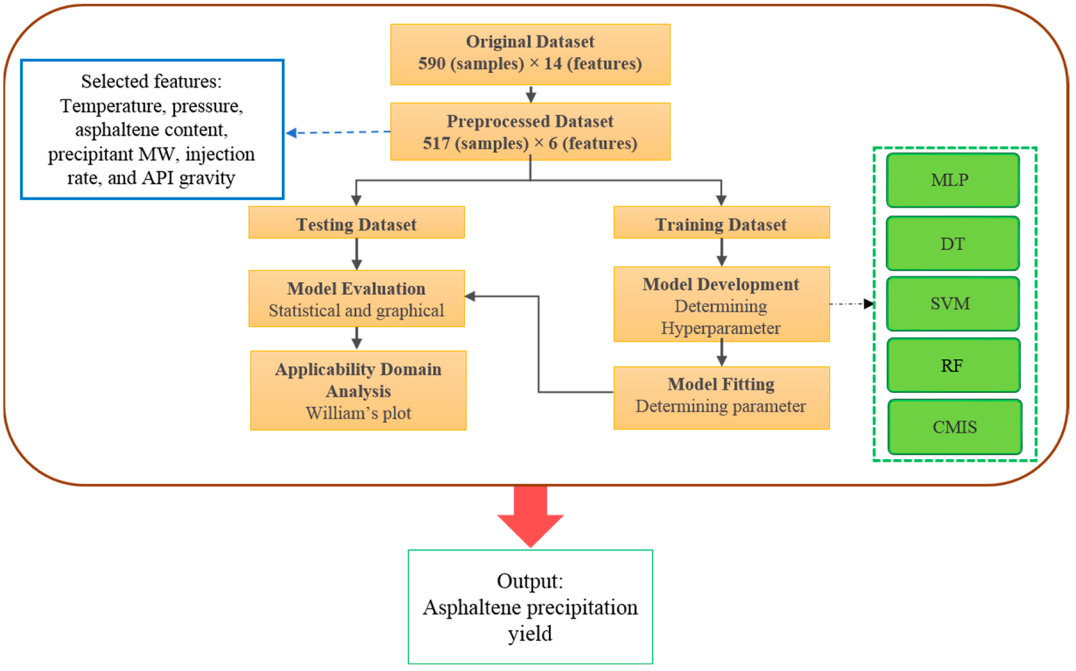


Figure 7. Flowchart of the modeling steps taken in the present research work.

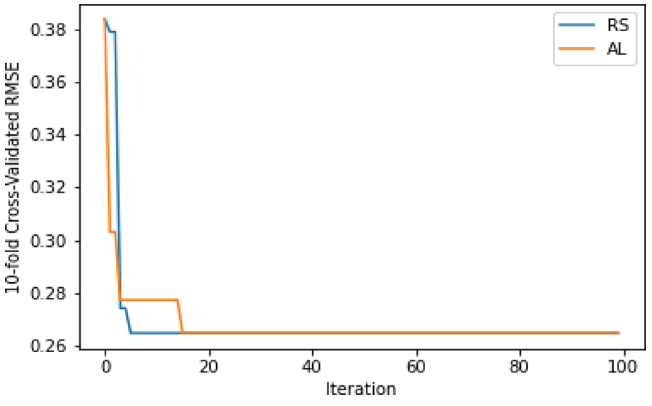


Figure 8. MLP error convergence results.

Table 3. MLP Optimization Results

	Range	RS	AL
hidden layer size	(1) to (15, 15)	(15, 13)	(15, 13)
best error	(1) to (15, 15)	0.2646	0.2646

3.2. Support Vector Machine. The support vector machine (SVM) model can be used for regression and classification problems. The distinctive characteristic of the SVM is the binary classifier. The SVM classifies the data by dividing it optimally into two groups.[40] The kernel coefficient (gamma) and the regularization parameter (C) are considered the most significant hyperparameters for the SVM optimization. The search space established for this instance is continuous. The results of the SVM modeling are demonstrated in Figure 9 and Table 4. RS exhibited the lowest error after 100 iterations for the best architecture, which is equal to 0.311.

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

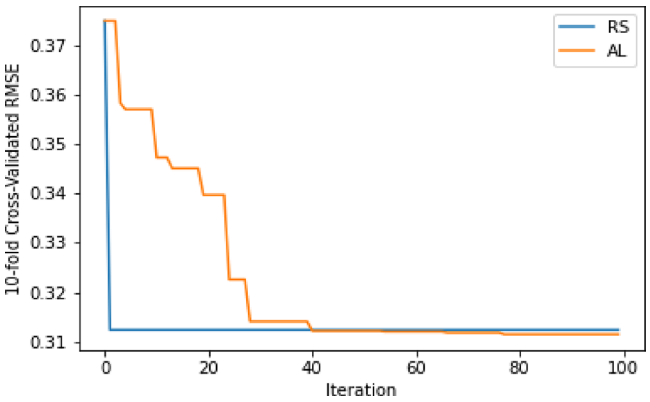


Figure 9. SVM error convergence results.

Table 4. SVM Optimization Results

	Range	RS	AL
C	0−1000	657.657	582.576
Gamma	0−1	0.088	0.092
Best error		0.312	0.311

3.3. Decision Tree. The decision tree (DT) model has a hierarchical architecture that resembles splitting into several branches. The input and decision rule define the partitioning of data at each node. The optimization process requires use of two significant DT hyperparameters that are the maximum depth and maximum number of features. DT modeling results are illustrated in Figure 10 and Table 5. RS has shown a far better error in comparison with the AL. The best error for the best architecture after 100 iterations is equal to 0.32.

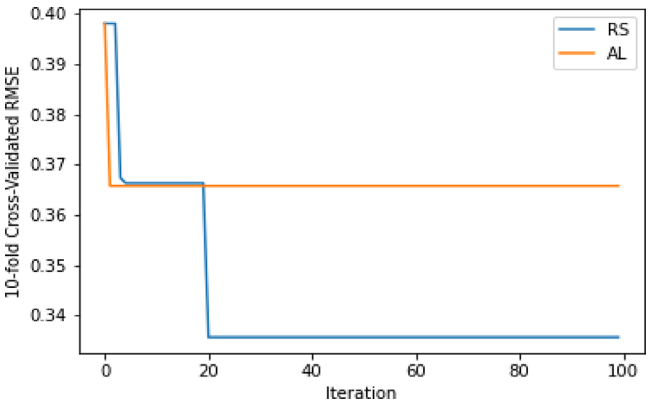


Figure 10. DT error convergence results.

Table 5. DT Optimization Results

	range	RS	AL
max depth	5−20	12	14
max features	1−6	4	3
best error		0.3356	0.3658

3.4. Random Forest. Numerous DTs can be combined into RF to obtain a more vigorous model in contrast to single separate DTs. A number of estimators are used to optimize the RF besides hyperparameters used for DT. The results of RF modeling are shown in Figure 11 and Table 6. The AL shows a better performance than the RL. After 100 iterations for the best structure, the best error is equal to 0.27.

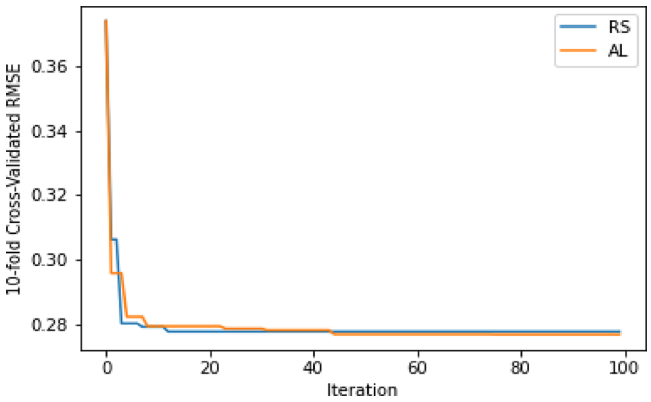


Figure 11. RF error convergence results.

Table 6. RF Optimization Results

	Range	RS	AL
Max depth	5−20	17	12
Max features	1−6	3	4
Number of estimators	5−100	45	75
Best error		0.2779	0.2771

3.5. Committee Machine Intelligent System. The committee machine intelligent system (CMIS) model can incorporate results of other models,[41] thereby producing more efficient solutions in comparison with individual solutions. This can be done using linear weighted averaging as shown in the following equation:

$$Y^i_{\mathrm{CMIS}} = a_1 \times Y^i_{\mathrm{MLP}} + a_2 \times Y^i_{\mathrm{SVM}} + a_0 \quad (2)$$

where, Y^i_{CMIS}, $Y^i_{\mathrm{MLP}}$and $Y^i_{\mathrm{SVM}}$are the predicted values by the CMIS, MLP, and SVM models, respectively and a_0, a_1, and a_2 are the constants. The best model from each modeling family with the original non-scaled data is used in the CMIS. The modeling results were obtained after 100 iterations, illustrated in Table 7. The most superior error is equal to 1.03.

Table 7. CMIS Results

	Coefficients		
Best error	a_1	a_2	a_0
1.03	0.931	0.074	−0.057

3.6. Population Balance Model. Nucleation and growth rate of asphaltene particles affect the asphaltene aggregation kinetics. The mechanism of asphaltene aggregation directly affects size of asphaltene particles.[21] Predicting the asphaltene growth rate and aggregation process helps to minimize the problems arising from asphaltene precipitation in flow assurance. The nuclei of asphaltene start to form under supersaturation conditions.[42] Then, these nuclei are attached together, and asphaltene aggregates will be formed. Following that, the collision between asphaltene aggregates causes to break large aggregates (death process) and produce smaller ones (birth process). This phenomenon continues until the birth and death process rates are equal and reach the equilibrium conditions. The PBM equations represent the number of particles in a specific size. The number of asphaltene aggregations based on the PBM is calculated as follows:

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

$$\frac{dC_k}{dt} = \frac{1}{2}K_{k-1,k-1}C_{k-1}^2 + C_{k-1}\sum_{j=1}^{k-2} K_{k-1,j}2^{j-k+1}C_j - C_k\sum_{j=1}^{k-1} K_{k,j}2^{j-k}C_j - C_k\sum_{j=k}^{N-1} K_{k,j}C_j \quad (3)$$

$$K_{i,j} = \frac{2}{3}\frac{RT}{\mu_m}\frac{(d_i + d_j)^2}{d_i d_j}\beta \quad (4)$$

where C ($\frac{\text{kmol}}{\text{m}^3}$) and t (min) are the asphaltene concentration and time, respectively. K is the kernel function and calculated using eq 4. i, j, and k are the number of particles. d is the diameter of particles. R and T are the gas constant value ($\frac{\text{J}}{\text{mol·K}}$) and temperature (K), respectively. β is an empirical constant. The distribution of aggregate is the output of the PBM. Hence, a mass transfer balance must be solved simultaneously to calculate the yield of asphaltene precipitation. Equations 5−7 are used to consider mass transfer between asphaltene particles and calculate the yield of asphaltene precipitation:

$$Y_{A,i}(t) = M_{A,i}(t)s_i = \frac{2^{i-1}M_{w,A}N_n s_i C_i(t)}{w_0} \quad (5)$$

$$s_i = \frac{\Delta x_i}{L} = \frac{r_c}{L}\left\{1 - \exp\left[-\frac{(\rho_A - \rho_m)d_i^{Df-1}\omega^2 t_c}{18\mu_m}\right]\right\} \quad (6)$$

$$Y_A(t) = \sum_{i=1}^{N} Y_{A,i}(t) \quad (7)$$

Δx_i is the distance between the aggregates. t_c and L are the centrifugation time and length of the centrifuge tube, respectively. r_c is the radius of the centrifuge, and ω is the angular velocity.

4. MODEL PERFORMANCE EVALUATION

In the following section, the accuracy of AI/ML models developed the present study are examined through the statistical error function, graphical error analysis, and leverage method.

4.1. Statistical Accuracy. In this research work, statistical tools including the root mean square error (RMSE), mean absolute relative deviation (MARD), mean relative deviation (MRD), coefficient of determination (R^2), and standard deviation were used to measure accuracy of the generated models. Equations 8−15 introduce these parameters as below:

$$D_i = X_{\text{Pred.}}^i - X_{\text{Exp.}}^i \quad (8)$$

$$RD_i = \frac{D_i}{|X_{\text{Pred.}}^i| + |X_{\text{Exp.}}^i|} \times 100 \quad (9)$$

$$\text{MRD} = \frac{1}{N}\sum_{i=1}^{N} RD_i \quad (10)$$

$$ARD_i = |RD_i| \quad (11)$$

$$\text{MARD} = \frac{1}{N}\sum_{i=1}^{N} |ARD_i| \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} D_i^2}{\sum_{i=1}^{N} (\bar{X}_{\text{Exp.}} - X_{\text{Pred.}}^i)^2}, \quad \bar{X}_{\text{Exp.}} = \frac{1}{N}\sum_{i=1}^{N} X_{\text{Exp.}}^i \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (D_i)^2} \quad (14)$$

$$\text{SD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (D_i - \bar{D}_i)^2} \quad (15)$$

According to the values for training, testing, and total datasets shown in Table 8, the best results are attained by the CMIS model. This is due to its nature of combining the benefits of several AI/ML models, therefore achieving a better result. The order of the performance for the created models is as follows: CMIS > SVM > MLP > RF > DT.

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

Model	Subset	RMSE	MARD	MRD	R^2	SD	N
MLP	training	1.08	15.65	2.40	0.97	1.08	413
	testing	1.73	18.90	0.50	0.94	1.72	104
	total	1.24	16.30	2.02	0.96	1.24	517
SVM	training	1.50	19.21	1.00	0.95	1.50	413
	testing	1.70	20.99	2.81	0.94	1.70	104
	total	1.54	19.57	1.36	0.95	1.54	517
DT	training	**0.40**	**6.13**	−3.33	**0.99**	**0.39**	413
	testing	3.67	20.98	−3.65	0.73	3.67	104
	total	1.68	**9.12**	−3.39	0.94	1.68	517
RF	training	0.79	11.82	8.82	0.99	0.79	413
	testing	2.40	**17.53**	9.32	0.88	2.40	104
	total	1.29	12.97	8.92	0.96	1.29	517
CMIS	training	1.08	15.66	**0.78**	0.97	1.08	413
	testing	**1.70**	18.36	**0.29**	**0.94**	**1.68**	104
	total	**1.23**	16.20	**0.68**	**0.96**	**1.23**	517

[a]The best values are in bold.

4.2. Graphical Error Analysis. The examination of MRD and RD_i values can enable comprehension of the deviation distribution. A comparison of the RD_i distribution for all the generated models is shown in Figure 12. A positive value of MRD signals over-prediction by the model. The more compact distribution around zero indicates superior performance of the model. The MRD value for the CMIS is 0.29% indicating the superior result for error distribution.

4.3. Leverage Method. The applicability domain analysis needs to be conducted to assess reliability of the developed models. The leverage method enables detection of unreliable outliers. The leverage known as hat values (h_i) can be identified using the Hat matrix (H):

$$H = X(X^{\text{T}}X)^{-1}X^{\text{T}} \quad (16)$$

The warning leverage (h^*) is used to determine the limits for data points in terms of hat values.

5. RESULTS AND DISCUSSION

In this section, the results obtained from the modeling are presented along with their accuracy assessment through error analysis, comparison with experimental data and the PBM results, analysis of the applicability domain, and processing time for the developed AI/ML models. In addition, the potential

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

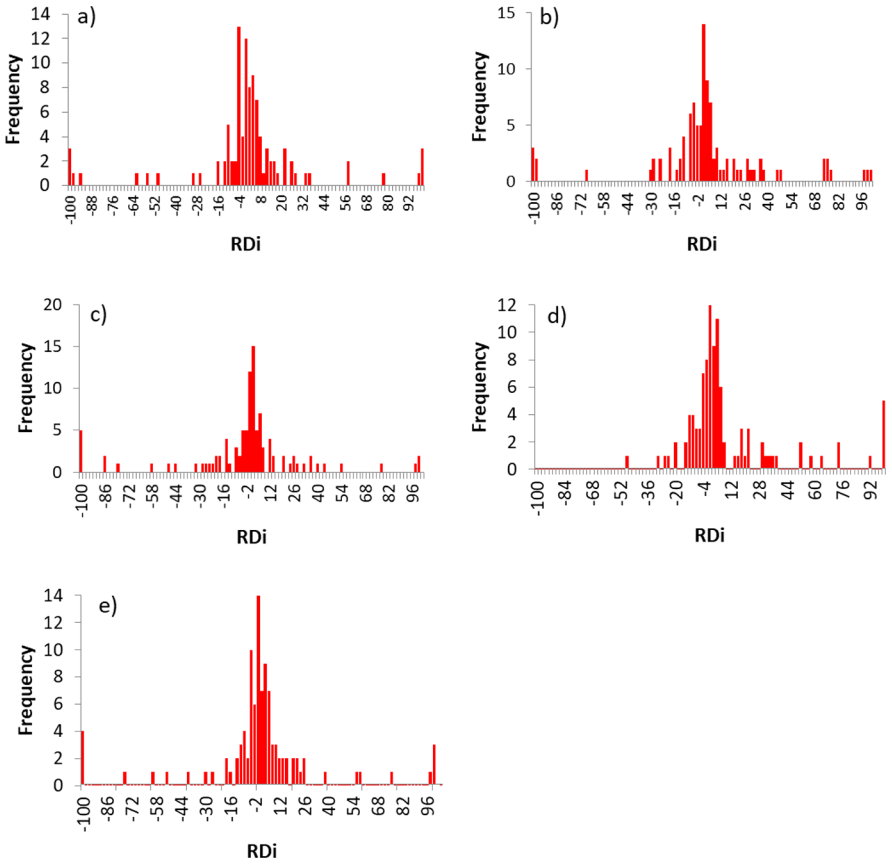


Figure 12. Relative deviation distribution for (a) MLP, (b) SVM, (c) DT, (d) RF, and (e) CMIS.

implications of the developed models are discussed. It is worth mentioning that the asphaltene precipitation yield (mass of asphaltene precipitation divided by the mass of bitumen in the feed) is considered as the target value in this research work to compare the experimental data and outcomes of the both PBM and AI/ML models.

5.1. Model Accuracy. The graphical accuracy charts displayed in this section illustrate the analysis of the CMIS model, which was determined as the best smart model among the developed models as mentioned previously. The cross-plot of the experimental and predicted asphaltene precipitation yield for training and testing datasets is presented in Figure 13. Based on Figure 13, most of the data samples for both training and testing datasets are compacted around the dashed bisector line, suggesting the CMIS model's superior performance. The model could accurately predict the asphaltene precipitation yield after training with the portion of the experimental data.

The relative deviation distribution based on experimental yield values for training and testing datasets is illustrated in Figure 14. Even though the majority of data points are dispersed between −20 and 20, for extremely low values of yield below 3%, relative deviation up to 100 was monitored in this plot. Hence, this model should not be used for asphaltene precipitation prediction for very small yields to prevent this kind of error. The low amount and low rate of injection of an anti-solvent to the system leads to a low yield of asphaltene precipitation. Hence,

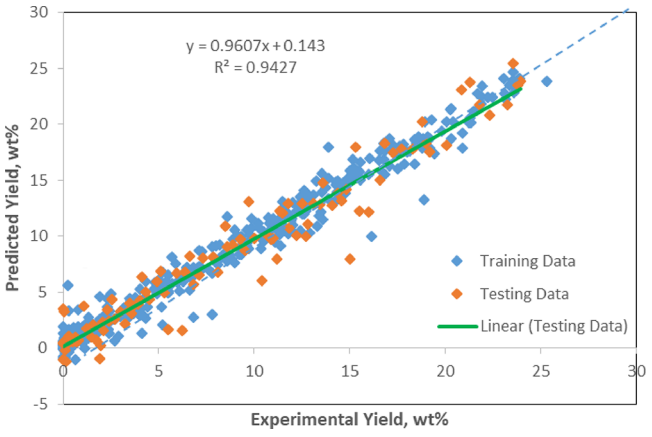


Figure 13. Cross plot of the experimental and predicted yield for training and testing datasets of the CMIS model.

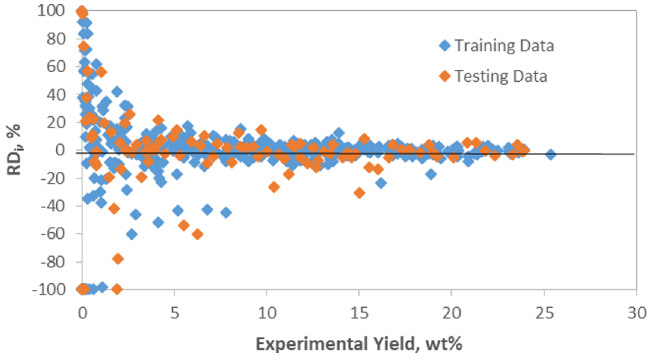
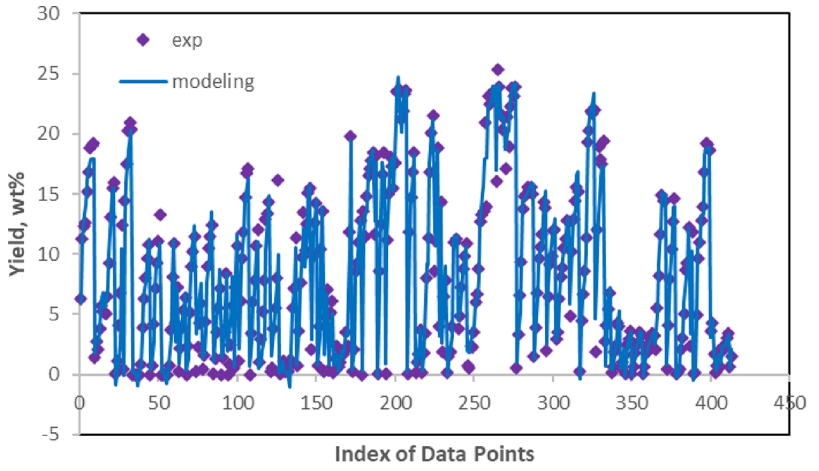
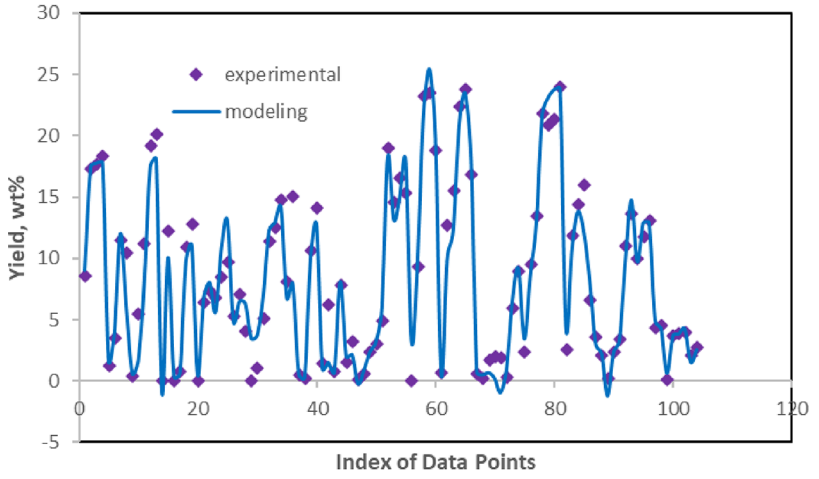


Figure 14. Relative deviation distribution plot for training and testing datasets of the CMIS model.

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137



a)



b)

Figure 15. Experimental and predicted yield for (a) training and (b) testing datasets.

estimating the accurate amount using experimental approaches in such conditions is challenging, causing errors in prediction. As a result, the experimental data for asphaltene precipitation with low yield is unreliable when using them in AI/ML methods.

Simultaneous depiction of experimental and predicted yield values for both training and testing datasets of the CMIS model is presented in Figure 15. The blue lines indicate the experimental yield values, and purple diamonds show the CMIS predictions. It can be concluded here that the CMIS prediction followed the experimental results of yield adequately, tightly.

5.2. Outlier Diagnosis. The measurement outliers for training and testing datasets of the CMIS model were examined using William's plot. It is evident from Figure 16 that the majority of data samples are situated in the safe zone that bounded between $SD_i = \pm 3$ and h^* (orange vertical line). This confirms the reliability of the obtained asphaltene precipitation yield results. Six training and five testing outliers were detected. Their details are outlined in Table 9. These are only low-leverage outliers present. The lack of high-leverage outliers can be attributed to data cleaning in the pre-processing stage.

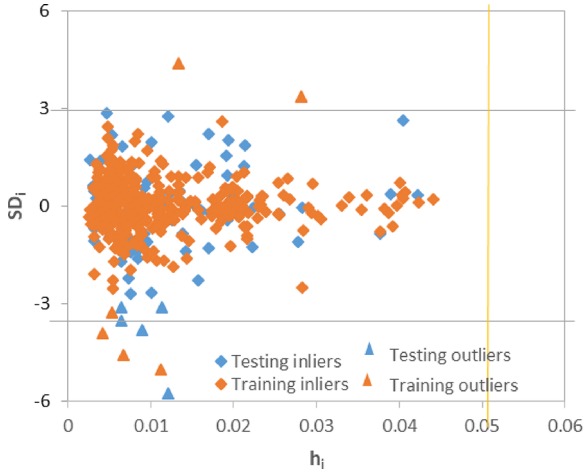


Figure 16. William's plot for the training and testing datasets of the CMIS model.

5.3. Comparison between the CMIS and the PBM. This section aims to compare the accuracy of the PBM and the

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

Table 9. Outlier Data Details

Index	Yield, wt %	Predicted yield, wt %	h_i	SD_i	Dataset	Leverage status
1	10.4	6.04	0.006	−3.55	testing	low
2	5.5	1.66	0.012	−3.14	testing	low
3	15.02	7.95	0.012	−5.77	testing	low
4	6.24	1.54	0.009	−3.83	testing	low
5	16.02	12.20	0.006	−3.11	testing	low
6	6.8	2.76	0.005	−3.29	training	low
7	7.8	2.99	0.004	−3.91	training	low
8	0.24	5.58	0.013	4.37	training	low
9	16.17	10	0.011	−5.03	training	low
10	13.91	17.99	0.028	3.37	training	low
11	18.89	13.27	0.006	−4.58	training	low

developed ML models in predicting the yield of asphaltene precipitation in the effect of *n*-alkane injection during bitumen recovery. Table 1 compares the experimental results (three bitumen samples), PBM, and CMIS outcomes in the present study. From data in Table 10, the performance of the PBM in predicting the yield of asphaltene precipitation (R^2 = 0.99) is higher than the CMIS model (R^2 = 0.94). However, the accuracy of the CMIS model is acceptable, and it is a good alternative for the experimental procedure. The CMIS model allows us to estimate the amount of asphaltene precipitation without costly experiment methods and/or solve complex mathematical relations. Definition of the correct initial and boundary conditions using experimental data is the main step in conducting the PBM. Lack of sensitive equipment and human errors prevent us from determining accurate initial/boundary conditions, and consequently, the chance of errors will increase. In addition, calculating the fractal dimension of asphaltene aggregates is complicated. Hence, most researchers assume that the asphaltene aggregates are spheres, which is a probable error source when solving population balance relations. To overcome these limitations, we need to spend high expenditure and time. Hence, these limitations encourage us to use smart models as alternatives for PBM in predicting target values in processes. Lack of data and variety of experimental samples are two main challenges during conducting ML models that cause deviation of ML results from actual data. Hence, the accuracy of the CMIS can be increased to the accuracy of population balance outcomes if these problems are solved.

Table 10. Comparison between Experimental Data, PBM, and CMIS Models Developed in this research work

	n-heptane wt %	Experimental, asphaltene yield wt %	PBM	CMIS
WC-B-B2	54.5	0.19	0.17	0.32
	57.5	2.12	2.12	2.18
	59.5	3.81	3.81	3.80
	64.5	5.82	5.99	5.73
MSB	38.5	0.01	0.03	0.15
	47.5	0.09	0.99	0.23
	50	0.12	0.12	0.26
	55	0.17	0.18	0.30
EU-HO-A1	80	0.008	0.02	0.15
	81.5	0.044	0.04	0.18
	83	0.109	0.14	0.24
	86	0.54	0.55	0.66
R^2			0.99	0.94

5.4. Processing Time. The processing time during conducting AI/ML models plays a vital role in the final cost. Considering the processing time can help to minimize the capital/operation cost. Faster processing time indicates better performance of the models by saving time. A comparison between run time and development time of modeling is illustrated in Figure 17. To facilitate the comparison, the time magnitudes were scaled in the range (0, 1). The maximum time resources were needed for the CMIS model in terms of both running and development phases. The SVM and MLP are inputs of the CMIS model, as mentioned before. Hence, the development time of the CMIS model is greater than the MLP and SVM, individually. Moreover, the DT model required the least time compared to other models. The MLP, SVM, and MLP had a longer run time, whereas SVM showed the need for longer development time.

5.5. Bitumen Recovery Implications. Developing reliable connectionist models can help to design and optimize deasphalting bitumen methods and manage unconventional oil production operating costs. The database constructed in this research work can also enhance the comprehension of the

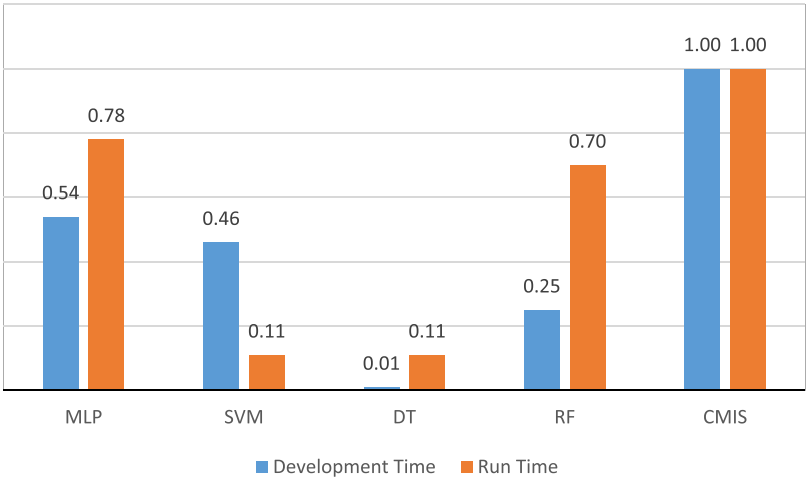


Figure 17. Time necessary to develop and run the models developed in this research work.

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

mechanism behind asphaltene precipitation upon introduction of precipitants.

6. CONCLUSIONS

The primary objective of this research work was to develop accurate predictive smarts models to estimate the yield of asphaltene precipitation during *n*-alkane injection for bitumen recovery. For this purpose, a database containing 590 data samples from 8 different experimental studies was gathered from the literature. This is so far the largest database assembled and reported up to now. Various pre-processing strategies such as removing duplicates and low-variance features, missing value imputation, collinearity assessment, description of data characteristics, outlier removal, feature selection, data splitting, and data scaling were used to prepare the data for modeling. Five ML models were constructed, including CMIS, MLP, SVM, DT, and RF to predict the asphaltene yield. Their accuracy was assessed using various error analysis methods. In addition, the results of the CMIS model are compared with the PBM. The following five main conclusions are drawn from this research work:

1. Three different categories of variables were obtained from the input features used including oil properties, precipitant properties, and operational parameters. The precipitant's API gravity, C_5 asphaltene content, MW, injection rate, pressure, and temperature are the most important features.
2. The results of outlier diagnosis revealed that six training and five testing outliers from the total dataset corresponded to 2.1% of all data samples used for modeling. This low amount of outlier calculation implies the generality of the CMIS model.
3. The CMIS model is determined as the best connectionist model in the present research work with an RMSE of 1.69 for the testing dataset. The R^2 for the testing dataset is equal to 0.94 indicating a good match between the experimental observations and the CMIS predictions. In addition, the MRD value of 0.29% for the testing dataset suggests that the error distribution for CMIS is symmetrical.
4. The connectionist models cannot predict the asphaltene precipitation yield accurately in low amount of asphaltene precipitation. The main reason for this problem is the low sensitivity of the experiment equipment to low asphaltene content that generates false data in this condition.
5. A comparison between the CMIS and the PBM results showed that the PBM is slightly more accurate than the CMIS. Nevertheless, performance of the CMIS model in predicting asphaltene yield is acceptable and it is a good alternative for the PBM considering the downsides of the PBM approach.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c03249.

Generated models in a convenient interface (ZIP)

Guidance 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

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

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

Afshin Tatar − *Petroleum Engineering Program, School of Mining and Geosciences, Nazarbayev University, Nur−Sultan 010000, Kazakhstan;* orcid.org/0000-0002-4571-4432

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c03249

Author Contributions

T. Yerkenov did the investigation, methodology, formal analysis, visualization, validation, and writing of the original draft. **S. Tazikeh** did the investigation, methodology, formal analysis, visualization, validation, and writing of the original draft. **A. Tatar** did the methodology, formal analysis, and visualization. **A. Shafiei** did the conceptualization, methodology, formal analysis, supervision, funding acquisition, project administration, resource gathering, data curation, writing of the review, and editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The financial support received from Nazarbayev University through a Collaborative Research Proposal (grant #091019CRP2103) is acknowledge and highly appreciated. The authors also would like to thank the editor and four anonymous reviewers for their timely and critical, yet fair and constructive comments that helped the authors to improve the quality and clarity of the manuscript.

NOMENCLATURE

C	regularization parameter
D_i	deviation
gamma	kernel coefficient
H	hat matrix
h^*	warning leverage
h_i	hat values
k	coefficient of proportionality
k	kurtosis value
N	number of data points
P	pressure
Q_1	first quartile
Q_3	third quartile
R	Spearman correlation coefficient
R^2	coefficient of determination
s	skewness value
T	temperature
w	asphaltene precipitation yield
X^T	transpose matrix X

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

Abbreviations

AI	artificial intelligence
AL	annealing
ANN	artificial neural network
API	American Petroleum Institute
ARD_i	absolute relative deviation
CMIS	committee machine intelligent system
DT	decision tree
EoS	equation of state
HPHT	high pressure, high temperature
IQR	interquartile range
MARD	mean absolute relative deviation
ML	machine learning
MLP	multilayer perceptron
MRD	mean relative deviation
MW	molecular weight
n-alkane	normal alkane
PBM	population balance model
RD	relative deviation
RF	random forest
RFE	recursive feature elimination
RMSE	root mean square error
RS	random search
SD, SD_i	standard deviation
SVM	support vector machine
VIF	variance inflation factor

Greek Letters

μ	oil viscosity
ν	precipitant to oil ratio
ρ	oil density
σ	solubility

Superscripts

T	transpose

Subscripts

crit	critical
solv	solvent

REFERENCES

(1) Duran, J. A. *Precipitation, Aggregation, and Settling of Asphaltenes from n-Alkane Diluted Bitumens*; University of Calgary, 2019, DOI: 10.11575/PRISM/37002.

(2) *Daily crude oil production in Canada from 1971 to 2019, by product type*; https://www.statista.com/statistics/485824/canadian-crudeoil-daily-production-by-type/, Access date 25/1/2022.

(3) Shafiee Neistanak, M. *Kinetics of asphaltene precipitation and flocculation from diluted bitumen*; Graduate Studies, 2014.

(4) Li, Z.; Firoozabadi, A. Modeling asphaltene precipitation by n-alkanes from heavy oils and bitumens using cubic-plus-association equation of state. *Energy Fuels* **2010**, *24*, 1106−1113.

(5) Nazari, F.; Assareh, M.; Asbaghi, E. V. Asphaltene formation modeling using vapor-liquid-liquid equilibrium calculations by PC-SAFT for reservoir and surface conditions. *J. Pet. Sci. Eng.* **2021**, *198*, No. 108209.

(6) Ramos-Pallares, F.; Santos, D.; Yarranton, H. W. Application of the Modified Regular Solution Model to Crude Oils Characterized from a Distillation Assay. *Energy Fuels* **2020**, *34*, 15270−15284.

(7) Yanes, J. F. R.; Feitosa, F. X.; do Carmo, F. R.; de Sant'Ana, H. B. Paraffin effects on the stability and precipitation of crude oil asphaltenes: Experimental onset determination and phase behavior approach. *Fluid Phase Equilib.* **2018**, *474*, 116−125.

(8) Ilyin, S. O.; Ignatenko, V. Y.; Kostyuk, A. V.; Levin, I. S.; Bondarenko, G. N. Deasphalting of heavy crude oil by hexamethyldisiloxane: The effect of a solvent/oil ratio on the structure, composition, and properties of precipitated asphaltenes. *J. Pet. Sci. Eng.* **2022**, *208*, No. 109329.

(9) Khormali, A.; Sharifov, A. R.; Torba, D. I. Experimental and modeling analysis of asphaltene precipitation in the near wellbore region of oil wells. *Pet. Sci. Technol.* **2018**, *36*, 1030−1036.

(10) Takahashi, S.; Hayashi, Y.; Takahashi, S.; Yazawa, N.; Sarma, H. *In Characteristics and impact of asphaltene precipitation during CO2 injection in sandstone and carbonate cores: An investigative analysis through laboratory tests and compositional simulation*, SPE International Improved Oil Recovery Conference in Asia Pacific, OnePetro: 2003.

(11) Brennan, L.; Owende, P. Biofuels from microalgae—A review of technologies for production, processing, and extractions of biofuels and co-products. *Renewable Sustainable Energy Rev.* **2010**, *14*, 557−577.

(12) Tharanivasan, A. K.; Yarranton, H. W.; Taylor, S. D. Application of a regular solution-based model to asphaltene precipitation from live oils. *Energy Fuels* **2011**, *25*, 528−538.

(13) Khormali, A.; Moghadasi, R.; Kazemzadeh, Y.; Struchkov, I. Development of a new chemical solvent package for increasing the asphaltene removal performance under static and dynamic conditions. *J. Pet. Sci. Eng.* **2021**, *206*, No. 109066.

(14) Ramos-Pallares, F.; Yarranton, H. W. Extending the Modified Regular Solution model to predict component partitioning to the asphaltene-rich phase. *Energy Fuels* **2020**, *34*, 5213−5230.

(15) Hoepfner, M. P.; Limsakoune, V.; Chuenmeechao, V.; Maqbool, T.; Fogler, H. S. A Fundamental Study of Asphaltene Deposition. *Energy Fuels* **2013**, *27*, 725−735.

(16) Haji-Akbari, N.; Masirisuk, P.; Hoepfner, M. P.; Fogler, H. S. A Unified Model for Aggregation of Asphaltenes. *Energy Fuels* **2013**, *27*, 2497−2505.

(17) Mousavi-Dehghani, S. A.; Riazi, M. R.; Vafaie-Sefti, M.; Mansoori, G. A. An analysis of methods for determination of onsets of asphaltene phase separations. *J. Pet. Sci. Eng.* **2004**, *42*, 145−156.

(18) Chaisoontornyotin, W.; Haji-Akbari, N.; Fogler, H. S.; Hoepfner, M. P. Combined Asphaltene Aggregation and Deposition Investigation. *Energy Fuels* **2016**, *30*, 1979−1986.

(19) Rashid, Z.; Wilfred, C. D.; Gnanasundaram, N.; Arunagiri, A.; Murugesan, T. A comprehensive review on the recent advances on the petroleum asphaltene aggregation. *J. Pet. Sci. Eng.* **2019**, *176*, 249−268.

(20) Subramanian, S.; Simon, S.; Sjöblom, J. Asphaltene Precipitation Models: A Review. *J. Dispersion Sci. Technol.* **2016**, *37*, 1027−1049.

(21) Khoshandam, A.; Alamdari, A. Kinetics of Asphaltene Precipitation in a Heptane−Toluene Mixture. *Energy Fuels* **2010**, *24*, 1917−1924.

(22) Moradi, S.; Hamed Mahvelati, E.; Ameli, F.; Dabir, B.; Rashtchian, D. Application of population balance equation in modeling of asphaltene particle size distribution and characterization of aggregation mechanisms under miscible gas Injection. *J. Mol. Liq.* **2017**, *232*, 207−213.

(23) Nassar, N. N.; Betancur, S.; Acevedo, S.; Franco, C. A.; Cortés, F. B. Development of a Population Balance Model to Describe the Influence of Shear and Nanoparticles on the Aggregation and Fragmentation of Asphaltene Aggregates. *Ind. Eng. Chem. Res.* **2015**, *54*, 8201−8211.

(24) Mohammed, I.; Mahmoud, M.; Al Shehri, D.; El-Husseiny, A.; Alade, O. Asphaltene precipitation and deposition: A critical review. *J. Pet. Sci. Eng.* **2021**, *197*, No. 107956.

(25) Sattari, M.; Kamari, A.; Mohammadi, A. H.; Ramjugernath, D. Rigorous model development for estimating asphaltene precipitation from crude oil by n-alkane titration. *Pet. Coal* **2019**, *61*, 998−1008.

(26) Bassir, S. M.; Madani, M. A new model for predicting asphaltene precipitation of diluted crude oil by implementing LSSVM-CSA algorithm. *Pet. Sci. Technol.* **2019**, *37*, 2252−2259.

(27) Ghorbani, M.; Zargar, G.; Jazayeri-Rad, H. Prediction of asphaltene precipitation using support vector regression tuned with genetic algorithms. *Petroleum* **2016**, *2*, 301−306.

(28) Ashoori, S.; Abedini, A.; Abedini, R.; Nasheghi, K. Q. Comparison of scaling equation with neural network model for prediction of asphaltene precipitation. *J. Pet. Sci. Eng.* **2010**, *72*, 186−194.

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137

(29) Hu, Y.-F.; Guo, T.-M. Effect of temperature and molecular weight of n-alkane precipitants on asphaltene precipitation. *Fluid Phase Equilib.* **2001**, *192*, 13−25.
(30) Johnston, K. Measurement and Modeling of Pentane-Diluted Bitumen Phase Behaviour; University of Calgary, 2017, DOI: 10.11575/PRISM/26844.
(31) Rivero Sanchez, J. A. Asphaltene Precipitation from Bitumen/Multicomponent Solvent Mixtures: University of Calgary, 2021, DOI: 10.11575/PRISM/38659.
(32) Luo, P.; Wang, X.; Gu, Y. Characterization of asphaltenes precipitated with three light alkanes under different experimental conditions. *Fluid Phase Equilib.* **2010**, *291*, 103−110.
(33) Davarpanah, L.; Vahabzadeh, F.; Dermanaki, A. Structural study of asphaltenes from Iranian heavy crude oil. *Oil Gas Sci. Technol.* **2015**, *70*, 1035−1049.
(34) Kokal, S. L.; Najman, J.; Sayegh, S. G.; George, A. E. Measurement and correlation of asphaltene precipitation from heavy oils by gas injection. *J. Can. Pet. Technol.* **1992**, *31*, PETSOC-92-04-01.
(35) Tatar, A.; Askarova, I.; Shafiei, A.; Rayhani, M. Data-Driven Connectionist Models for Performance Prediction of Low Salinity Waterflooding in Sandstone Reservoirs. *ACS Omega* **2021**, *6*, 32304−32326.
(36) Chiu, P. C.; Selamat, A.; Krejcar, O.; Kuok, K. K.; Bujang, S. D. A.; Fujita, H. Missing Value Imputation Designs and Methods of Nature-Inspired Metaheuristic Techniques: A Systematic Review. *IEEE Access* **2022**, *10*, 61544−61566.
(37) Bilogur, A. Missingno: a missing data visualization suite. *J. Open Source Software* **2018**, *3*, 547.
(38) Rashid, S.; Ghamartale, A.; Abbasi, J.; Darvish, H.; Tatar, A. Prediction of Critical Multiphase Flow Through Chokes by Using A Rigorous Artificial Neural Network Method. *Flow Meas. Instrum.* **2019**, *69*, No. 101579.
(39) Orhan, U.; Hekim, M.; Ozer, M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Syst. Appl.* **2011**, *38*, 13475−13481.
(40) Shafiei, A.; Parsaei, H.; Dusseault, M. B. Rock Squeezing Prediction By a Support Vector Machine Classifier. In *46th U.S. Rock Mechanics/Geomechanics Symposium*; 2012; pp ARMA-2012-435
(41) Nilsson, N.; Machines, L. Foundations of trainable pattern classifying systems. McGraw-Hill, New York OBrien RM (2007) A caution regarding rules of thumb for variance ination factors. *Qual Quant.* **1965**, *41*, 673.
(42) Solaimany-Nazar, A. R.; Rahimi, H. Dynamic Determination of Asphaltene Aggregate Size Distribution in Shear Induced Organic Solvents. *Energy Fuels* **2008**, *22*, 3435−3442.

https://doi.org/10.1021/acsomega.2c03249
ACS Omega 2022, 7, 33123−33137