ACS OMEGA

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

Article

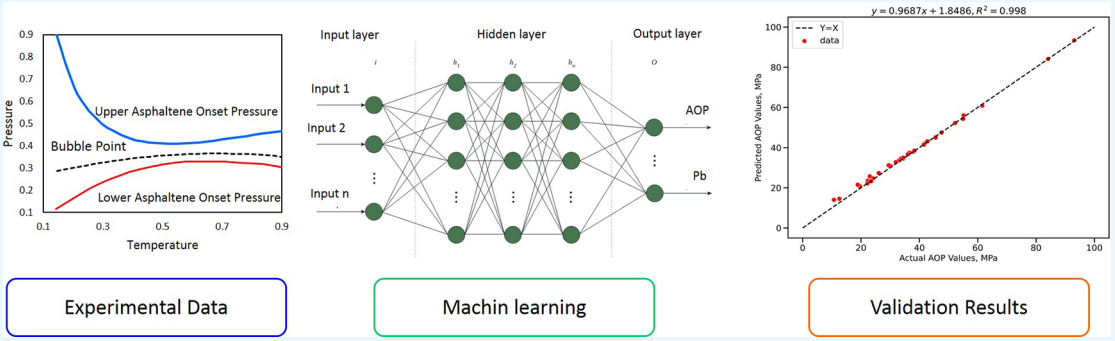
# A Comparison between the Perturbed-Chain Statistical Associating Fluid Theory Equation of State and Machine Learning Modeling Approaches in Asphaltene Onset Pressure and Bubble Point Pressure Prediction during Gas Injection

Simin Tazikeh, Abdollah Davoudi, Ali Shafiei,* Hossein Parsaei, Timur Sh. Atabaev, and Oleksandr P. Ivakhnenko

Cite This: *ACS Omega* 2022, 7, 30113−30124

Read Online

ACCESS | Metrics & More | Article Recommendations




**ABSTRACT:** Predicting asphaltene onset pressure (AOP) and bubble point pressure (Pb) is essential for optimization of gas injection for enhanced oil recovery. Pressure-Volume-Temperature or PVT studies along with equations of state (EoSs) are widely used to predict AOP and Pb. However, PVT experiments are costly and time-consuming. The perturbed-chain statistical associating fluid theory or PC-SAFT is a sophisticated EoS used for prediction of the AOP and Pb. However, this method is computationally complex and has high data requirements. Hence, developing precise and reliable smart models for prediction of the AOP and Pb is inevitable. In this paper, we used machine learning (ML) methods to develop predictive tools for the estimation of the AOP and Pb using experimental data (AOP data set: 170 samples; Pb data set: 146 samples). Extra trees (ET), support vector machine (SVM), decision tree, and k-nearest neighbors ML methods were used. Reservoir temperature, reservoir pressure, SARA fraction, API gravity, gas−oil ratio, fluid molecular weight, monophasic composition, and composition of gas injection are considered as input data. The ET ($R^2$: 0.793, RMSE: 7.5) and the SVM models ($R^2$: 0.988, RMSE: 0.76) attained more reliable results for estimation of the AOP and Pb, respectively. Generally, the accuracy of the PC-SAFT model is higher than that of the AI/ML models. However, our results confirm that the AI/ML approach is an acceptable alternative for the PC-SAFT model when we face lack of data and/or complex mathematical equations. The developed smart models are accurate and fast and produce reliable results with lower data requirements.

## 1. INTRODUCTION

Flow assurance issues are well known in the petroleum industry spanning from upstream to downstream. Asphaltene precipitation and deposition is one of the main issues in this industry that occurs due to changes in temperature, pressure, and oil composition. Asphaltenes are crude oil's polar components with a complex molecular structure and different molecular weights (from 500 to 10,000 g/mol).[1] Based on their solubility, asphaltenes dissolve in aromatic solvents (e.g., toluene, naphtha, and benzene) and are nonsoluble in *n*-alkanes. Clogging porous media, well plugging, permeability reduction, and wettability alteration are some of the significant consequences of asphaltene precipitation, which occur during different oil production stages.[2,3]

Injection of gas (e.g., nitrogen, carbon dioxide, and natural gas) during EOR operations is one of the widely used and effective ways to recover residual oil after primary and

Received: May 22, 2022
Accepted: August 1, 2022
Published: August 16, 2022

secondary oil recovery methods. The injection gases can enhance the flow of crude oil via reducing its viscosity.[3] However, a combination of oil and gas changes the composition and behavior of reservoir fluids and provides a favorable situation for asphaltene precipitation.[2] Hence, estimation of asphaltene onset pressure (AOP) is vital to minimize asphaltene precipitation and the related problems.

Accurate estimation of the AOP in various stages of oil production can help to have a correct long-term plan from the economic viewpoint. Various research studies (experimental and modeling) addressed the main problems in the petroleum industry in the field of kinetics, AOP, and the amount of asphaltene precipitation/deposition. Light scattering, high-pressure microscopy, gravimetric precipitation, quartz crystal resonators, and isobaric filtration techniques are the main experimental methods that are used for estimating the AOP.[4−6] Moreover, there are many modeling approaches to assess the kinetic and asphaltene thermodynamic states under various thermodynamic conditions.

Thermodynamic solubility techniques and colloidal-based models are two main groups of modeling asphaltene precipitation considering the assumption about the stability state of asphaltene in crude oil.[7,8] In the colloidal approach, asphaltene is assumed as a solid particle peptized by resin in the solution. Based on this theory, the asphaltene aggregation will be formed when resins separate from asphaltene molecules. In addition, the asphaltene precipitation assumes an irreversible phenomenon in the colloidal theory. In the solubility model, asphaltenes are considered soluble particles in crude oil. In addition, this theory assumes that the dispersion force between asphaltene and other fractions in crude oil is responsible for asphaltene aggregation, not polar interaction. Asphaltenes will precipitate by decreasing their concentration below a critical solubility value.[9]

The solubility models fall into two main categories of regular solution theory and equation of state (EOS). The first group considers that the system is made of crude oil and asphaltene. The regular solution theory is used generally for polymer solution and studies based on Flory−Huggins,[10] Scatchard−Hildebrand,[11] and Scott−Magat theories.[12] EOS theory is considered as a mixture of pseudo- and pure components.[9] Cubic, Cubic Plus Association, statistical associating fluid theory (SAFT), Peng Robinson, and Soave−Redlich−Kwong are placed in EOS theory.[7,9] Among these, the SAFT model is the most reliable model to investigate thermodynamic behavior of asphaltene as a complex mixture because of the association effect, and nonspherical chains are considered in this theory.[13] The perturbed-chain form statistical associating fluid theory (PC-SAFT) is the most useful SAFT model to predict the stability of asphaltene during EOR methods (e.g., gas injection), drop of pressure, and PVT studies that were proposed and developed by Gross and Sadowski.[14]

Several research studies applied the SAFT model to simulate asphaltene stability in the system. Ting et al.[13] used the SAFT model for this aim. They classified crude oil components into three gas pseudocomponents and three liquid phases. Their model successfully predicted the instability and bubble point of asphaltene compared to experimental data. Following that, Panuganti et al.[15] improved the parameters used by Ting et al.[13] and considered the effect of gas injection on the stability of asphaltene in the system. They reported that considering light components as a separate group, the model's accuracy in predicting bubble point and AOP was enhanced. Gonzalez et al.[16] investigated the effect of signal component gas (e.g., $CH_4$, $N_2$, and $CO_2$) on asphaltene stability. In another study, they used the PC-SAFT model for accessing asphaltene precipitation in the presence of oil-based mud and considering the gas-oil-ratio (GOR).[17] In addition to results of these studies, the thermodynamic modeling approach is costly, time-consuming, usually complicated, and unsuitable for nonlinear systems. Hence, it is necessary to use methods that cover this limitation because asphaltene precipitation and deposition is a nonlinear and complex process.

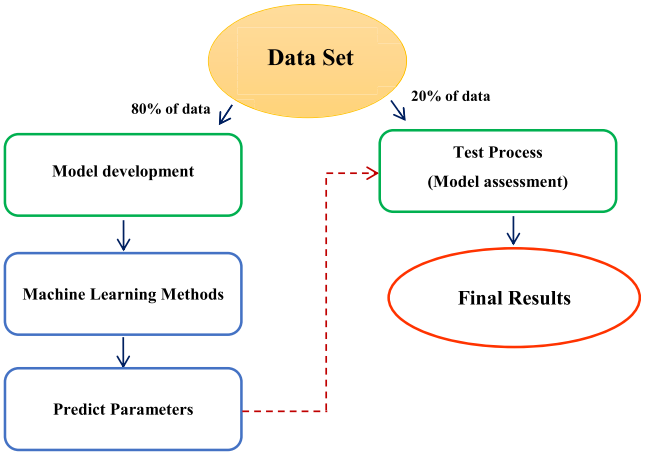



**Figure 1.** Flowchart showing the AI/ML modeling workflow used in this research work for prediction of AOP and Pb.

**Table 1. Number and Percentage of Missing Values for Each Parameter**

| parameter | number of missing data | percentage of missing data (%) |
|---|---|---|
| reservoir pressure | 50 | 29.41 |
| reservoir temperature | 31 | 18.23 |
| API gravity | 0 | 0 |
| GOR | 14 | 8.23 |
| fluid molecular weight | 5 | 2.94 |
| oil-$N_2$ | 0 | 0 |
| oil-$CO_2$ | 0 | 0 |
| oil-$H_2S$ | 0 | 0 |
| oil-$C_1$ | 0 | 0 |
| oil-$C_2$ | 0 | 0 |
| oil-$C_3$ | 0 | 0 |
| oil-i-*n*-$C_4$ | 0 | 0 |
| oil-i-*n*-$C_5$ | 0 | 0 |
| oil-$C_6$ | 0 | 0 |
| saturates | 5 | 2.94 |
| aromatics | 5 | 2.94 |
| resins | 5 | 2.94 |
| asphaltenes | 5 | 2.94 |
| $N_2$ | 0 | 0 |
| $CO_2$ | 0 | 0 |
| $H_2S$ | 0 | 0 |
| $C_1$ | 0 | 0 |
| $C_2$ | 0 | 0 |
| $C_3$ | 0 | 0 |
| $C_{4+}$ | 0 | 0 |
| temperature | 0 | 0 |
| AOP | 0 | 0 |
| bubble point pressure (Pb) | 24 | 14.12 |



https://doi.org/10.1021/acsomega.2c03192
ACS Omega 2022, 7, 30113−30124

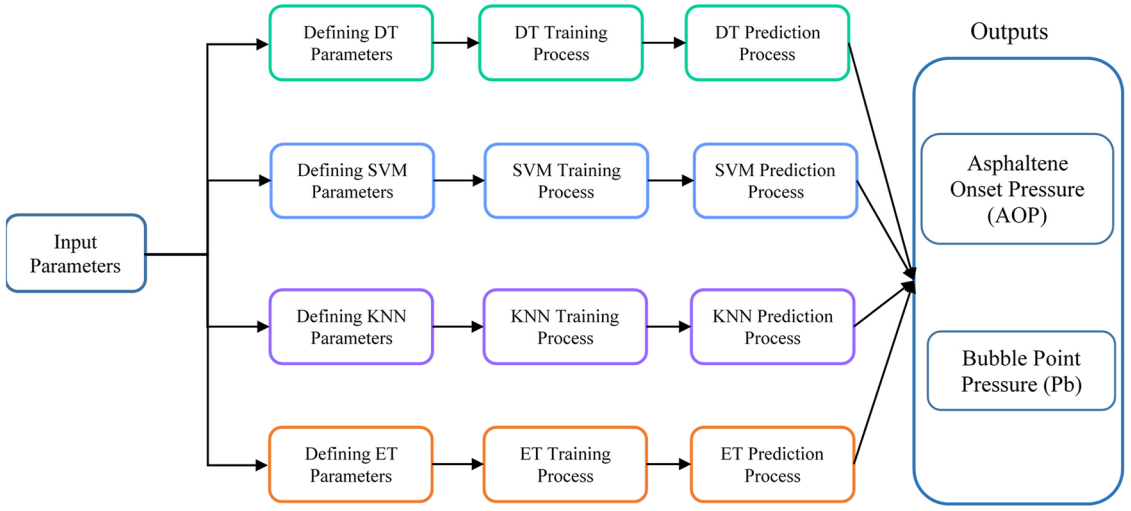



**Figure 2.** Overall schematic flowchart for the used models.

Smart computational tools are considered practical methods to predict the target parameters without knowing the mathematical/empirical relations and experimental procedure. Machine learning (ML) methods, which are data-driven models, attempt to find a function relating predictors to output.[18] This approach provides a suitable method to predict asphaltene thermodynamic behavior as a nonlinear and discontinuous function. Artificial neural network (ANN), least-squares support vector machine (LSSVM), support vector machine (SVM) and support vector regression (SVR), random forest (RF), genetic algorithm (GA), decision tree (DT), and particle swarm optimization (PSO) are the most common AL algorithms used for the prediction of asphaltene stability parameters. Khamehchi et al.[19] used ANN and adaptive neural fuzzy system models to predict bubble point and AOP during $CO_2$ injection. Their results are in good agreement with experimental data and confirm validity of their models. Zendehboudi et al.[18] investigated the effect of miscible gas injection during EOR using the PSO-ANN method. Based on their results, temperature has a great impact on minimum miscibility pressure (MMP), and a good agreement between their modeling and experimental results was observed. Kamari et al.[20] predicted AOP and oil saturation conditions using the LSSVM. They considered different crude oils mostly from the Middle East with different asphaltene contents. Their results were by comparing the results with experimental data. In another study, LSSVM and a multilayer extreme learning machine were used to estimate asphaltene bubble pressure by Rashidi et al.[21] They used the GA and PSO to optimize the methods. In addition to the accuracy of their models in predicting bubble point and oil formation volume factor, the results show that gas−oil gravity is an influential factor in bubble point prediction compared with other parameters in their work. Although the results of previous studies are acceptable, the effects of various types of gas injection on AOP and bubble point and the relation between ML results and the PC-SAFT model were not considered in previous research studies reported in the literature.

The main objective of the present study was to use AI/ML methods to estimate bubble point pressure and AOP and compare the results with the PC-SAFT model during gas injection. To the best of our knowledge, this is the first research work where the outcomes of the PC-SAFT model and AI/ML-based smart models in predicting AOP and bubble point pressure during gas injection are compared.

This paper is organized into four main parts. After Introduction, methodology is described in detail, and data preprocessing, model development, and evaluation with the assessment of statistical error and PC-SAFT EoS are presented. Subsequently, in the Results and Discussion section, modeling results, evaluation of the developed models, and comparisons between the developed smart models with PC-SAFT EoS are discussed. Finally, the main outcomes are highlighted in the Conclusions section.

## 2. METHODOLOGY

The overall procedure for this modeling research work consists of three steps: data collection, data preprocessing, AI/ML model development, and model evaluation. Each step is discussed in detail in the following subsections.

**2.1. Data Collection.** Various parameters affect AOP and bubble point pressure during gas injection. These parameters include temperature, pressure, molecular weight of asphaltene, oil composition, oil properties (e.g., API and SARA), GOR, and properties of gas injection. For the purpose of this research work, we used experimental data from the literature where the effect of these parameters on AOP and Pb was investigated by several researchers.[15,22−43] The collected data were then split into two sets of training data that include 80% of the data set (AOP data set: 136 samples; Pb data set: 116 samples) and testing data that contain 20% (AOP data set: 34 samples; Pb data set: 30 samples) of the total data. Training data are used to train and build AI/ML models, and the testing data are used to evaluate the performance of the developed smart models in estimating the desired parameter. A flowchart illustrating the AI/ML modeling approach for prediction of AOP and Pb is presented in Figure 1.

**2.2. Data Preprocessing.** *2.2.1. Missing Value Imputation.* The issue of missing values in data-driven modeling approaches is quite common when analyzing real data sets. Missing values can cause a bias in modeling results and affect model performance. Hence, a key step in data modeling is handling missing value. In this research work, AOP and Pb are target variables. Hence, samples with missing values in these two parameters were removed from the database. The number of missing values in each parameter must first be specified to



https://doi.org/10.1021/acsomega.2c03192
ACS Omega 2022, 7, 30113−30124

Table 2. Statistical Information of the Data Set for AOP

| parameter | count | unit | range | mean (standard deviation) |
|---|---|---|---|---|
| reservoir pressure | 170 | MPa | 19.68−137.90 | 42.01 (16.16) |
| reservoir temperature | 170 | K | 294.25−418.87 | 383.13 (21.17) |
| API gravity | 170 | | 20.32−41.69 | 34.37 (5.84) |
| GOR | 170 | $m^3/m^3$ | 38.12−314.78 | 142.39 (66.02) |
| fluid molecular weight | 170 | g/mol | 66.69−169.00 | 108.37 (26.56) |
| oil-$N_2$ | 170 | mol % | 0.02−6.59 | 0.52 (1.06) |
| oil-$CO_2$ | 170 | mol % | 0.05−3.68 | 1.54 (0.97) |
| oil-$H_2S$ | 170 | mol % | 0.00−5.39 | 0.44 (1.26) |
| oil-$C_1$ | 170 | mol % | 12.39−61.00 | 34.11 (11.52) |
| oil-$C_2$ | 170 | mol % | 4.33−12.42 | 7.59 (2.01) |
| oil-$C_3$ | 170 | mol % | 2.48−10.29 | 6.18 (1.29) |
| oil-i-*n*-$C_4$ | 170 | mol % | 0.00−6.90 | 4.38 (2.02) |
| oil-i-*n*-$C_5$ | 170 | mol % | 0.00−7.85 | 3.77 (1.55) |
| oil-$C_6$ | 170 | mol % | 0.00−5.48 | 2.95 (1.48) |
| saturates | 170 | wt % | 24.80−80.64 | 61.43 (12.39) |
| aromatics | 170 | wt % | 11.60−47.81 | 26.84 (8.22) |
| resins | 170 | wt % | 1.47−20.60 | 8.36 (4.57) |
| asphaltenes | 170 | wt % | 0.10−15.50 | 2.88 (4.61) |
| $N_2$ | 170 | mol % | 0.00−10.00 | 0.34 (1.61) |
| $CO_2$ | 170 | mol % | 0.00−45.00 | 4.07 (9.62) |
| $H_2S$ | 170 | mol % | 0.00−10.00 | 0.22 (1.33) |
| $C_1$ | 170 | mol % | 0.00−46.52 | 5.00 (10.09) |
| $C_2$ | 170 | mol % | 0.00−7.75 | 0.78 (1.62) |
| $C_3$ | 170 | mol % | 0.00−5.94 | 0.44 (1.00) |
| $C_{4+}$ | 170 | mol % | 0.00−5.21 | 0.29 (0.72) |
| temperature | 170 | K | 294.26−424.55 | 362.62 (30.56) |
| asphaltene onset pressure (AOP) | 170 | MPa | 10.71−135.14 | 41.26 (19.99) |

Table 3. Statistical Information of the Data Set for Pb

| parameter | count | unit | range | mean (standard deviation) |
|---|---|---|---|---|
| reservoir pressure | 146 | MPa | 19.68−89.63 | 41.46 (12.44) |
| reservoir temperature | 146 | K | 294.25−418.87 | 384.09 (22.26) |
| API gravity | 146 | | 20.32−41.69 | 34.87 (5.99) |
| GOR | 146 | $m^3/m^3$ | 56.81−314.78 | 152.63 (62.20) |
| fluid molecular weight | 146 | g/mol | 66.69−169.00 | 103.43 (21.77) |
| oil-$N_2$ | 146 | mol % | 0.02−6.59 | 0.58 (1.13) |
| oil-$CO_2$ | 146 | mol % | 0.05−3.68 | 1.65 (0.99) |
| oil-$H_2S$ | 146 | mol % | 0.00−5.39 | 0.51 (1.35) |
| oil-$C_1$ | 146 | mol % | 16.84−61.00 | 35.62 (10.06) |
| oil-$C_2$ | 146 | mol % | 4.33−12.42 | 7.78 (1.92) |
| oil-C3 | 146 | mol % | 2.48−10.29 | 6.21 (1.32) |
| oil-i-*n*-$C_4$ | 146 | mol % | 0.00−6.90 | 4.37 (2.04) |
| oil-i-*n*-$C_5$ | 146 | mol % | 0.00−7.85 | 3.74 (1.52) |
| oil-$C_6$ | 146 | mol % | 0.00−5.48 | 2.89 (1.43) |
| saturates | 146 | wt % | 24.80−80.64 | 61.99 (13.06) |
| aromatics | 146 | wt % | 11.60−47.81 | 27.14 (8.58) |
| resins | 146 | wt % | 1.47−20.60 | 7.66 (4.32) |
| asphaltenes | 146 | wt % | 0.10−15.50 | 2.88 (4.77) |
| $N_2$ | 146 | mol % | 0.00−10.00 | 0.29 (1.63) |
| $CO_2$ | 146 | mol % | 0.00−45.00 | 3.73 (9.35) |
| $H_2S$ | 146 | mol % | 0.00−10.00 | 0.23 (1.42) |
| $C_1$ | 146 | mol % | 0.00−39.46 | 3.23 (7.47) |
| $C_2$ | 146 | mol % | 0.00−6.79 | 0.45 (1.07) |
| $C_3$ | 146 | mol % | 0.00−5.15 | 0.26 (0.72) |
| $C_{4+}$ | 146 | mol % | 0.00−5.21 | 0.23 (0.70) |
| temperature | 146 | K | 294.26−418.71 | 362.37 (31.35) |
| bubble point pressure (Pb) | 146 | MPa | 5.38−43.21 | (7.17) |

deal with the remaining missing values. A parameter must be eliminated if the parameter has many missing values or the percentage of missing values exceeds a certain threshold. In this research work, a threshold of 50% was considered according to previous research studies.[44] As shown in Table 1, no parameter has more than 50% missing values, and therefore, they remain. For the other parameters, the missing values are replaced with the mean value of the parameter which is a widely used method for estimating missing values for numeric features. There are many approaches available for replacing missing values with imputed values, and every single of them has its own pros and cons. A detailed discussion of those methods is beyond the scope of this research work. In this research work, one of the most used and popular methods was used, which is mean (average) imputation as a quick and simple way for handling missing values, and in this case, fair results have been achieved.[45]

*2.2.2. Data Normalization.* Data normalization was used to normalize the range of variables such that the range of features is the same. This process is commonly performed during the data preprocessing step to handle highly varying magnitudes or values or units. All input parameters have the same effect on the model development process, ultimately. In this research work, the standardization technique was used for data normalization that causes all parameters have a mean and standard deviation of 0 and 1, respectively. A standard scalar is first fitted on the training data and then applied to the testing data.

*2.2.3. Model Development.* Four AI/ML models that have been shown to be effective in data modeling problems were explored in this research work: SVM, KNN, and DT, and ET. A brief description of each AI/ML algorithm is provided in the text below.

*2.2.4. Support Vector Machine.* The SVM is a powerful tool in regression and classification that was introduced in the 1960s.[45] The training data are classified in different classes in this method and show a good performance in modeling noisy and high-dimension data. In addition, it is suitable for a database containing a high number of features compared to the number of samples.[46]

*2.2.5. k-Nearest Neighbor.* The KNN method predicts a class label for a given sample based on its K neighborhods.[47] The KNN algorithm works by finding the K examples closest to the query first and then averages the labels of these K examples to estimate the value for the query. We used the KNN algorithm because it is easy to use and effective and robust in data modeling.

*2.2.6. Decision Tree.* The DT algorithm provides a rule-based model by which the value of a target variable is estimated using decision rules extracted from the data features.[48] The extracted tree can be considered as a piecewise constant approximation. For the sake of briefness, the DT algorithm is not discussed in detail; such information is provided elsewhere.

*2.2.7. Extra Trees.* The ET algorithm is a type of ensemble learning technique in which the results of different de-correlated DT algorithms are aggregated to provide output



https://doi.org/10.1021/acsomega.2c03192
ACS Omega 2022, 7, 30113−30124

**Table 4. Result of the AI/ML Model Development for Prediction of AOP**

| model | subset | $R^2$ (STD) | RMSE (STD) |
|---|---|---|---|
| extra trees | training | 1.00 (0.00) | $6.72 \times 10^{-14}$ (0.00) |
| | testing | **0.79 ($1.11 \times 10^{-16}$)** | **7.51 ($1.78 \times 10^{-15}$)** |
| k-nearest neighbors | training | 1.00 (0.00) | 0.00 (0.00) |
| | testing | 0.62 ($2.22 \times 10^{-16}$) | 10.16 (0.00) |
| decision tree | training | 0.71 ($2.22 \times 10^{-16}$) | $1.10 \times 10^{1}$ ($1.78 \times 10^{-15}$) |
| | testing | 0.48 ($1.10 \times 10^{-16}$) | 11.87 ($1.78 \times 10^{-15}$) |
| support vector machine | training | 0.97 ($3.33 \times 10^{-16}$) | 3.61 ($4.44 \times 10^{-16}$) |
| | testing | 0.76 ($2.22 \times 10^{-16}$) | 7.99 ($8.88 \times 10^{-16}$) |

**Table 5. Result of AI/ML Model Development for Prediction of Pb**

| model | subset | $R^2$ (STD) | RMSE (STD) |
|---|---|---|---|
| extra trees | training | 1.00 (0.00) | $2.97 \times 10^{-14}$ (0.00) |
| | testing | 0.97 ($2.22 \times 10^{-16}$) | 1.71 ($6.66 \times 10^{-16}$) |
| k-nearest neighbors | training | 1.00 (0.00) | 0.00 (0.00) |
| | testing | 0.94 (0.00) | 1.84 ($6.66 \times 10^{-16}$) |
| decision tree | training | 0.99 (0.00) | $2.54 \times 10^{-01}$ (0.00) |
| | testing | 0.88 ($3.33 \times 10^{-16}$) | 2.51 ($4.44 \times 10^{-16}$) |
| support vector machine | training | 0.98 ($3.33 \times 10^{-16}$) | $9.36 \times 10^{-01}$ ($1.11 \times 10^{-16}$) |
| | testing | **0.99 (0.00)** | **0.79 (0.00)** |

of the model. In the ET method, a group of trees are formed randomly, and their results are aggregated.[49] The overall flowchart of all models is shown in Figure 2.

Each AI/ML model contains several essential parameters known as hyperparameters that significantly influence the model performance in predicting the output parameter. It should be emphasized here that hyperparameter optimization must be completed before training the models.[50,51] In this research work, an automatic hyperparameter optimization framework specifically intended for ML known as Optuna[52] is used in conjunction with fivefold cross-validation on training data to optimize hyperparameters. To accomplish this, we first determine the maximum and lowest values of each hyperparameter, and then Optuna creates a model by arranging these parameters in a random order. Then performance of the model is evaluated using 5-fold cross-validation. As a result, the model will be constructed in such a way that the root mean square error (RMSE) is minimized. This method is repeated until the RMSE is as low as possible.[52]

*2.2.8. Model Evaluation.* We used two well-known indices, namely, coefficient of determination ($R^2$) and RMSE as defined below to evaluate the performance of the developed models:[53]

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - y)^2} \quad (1)$$

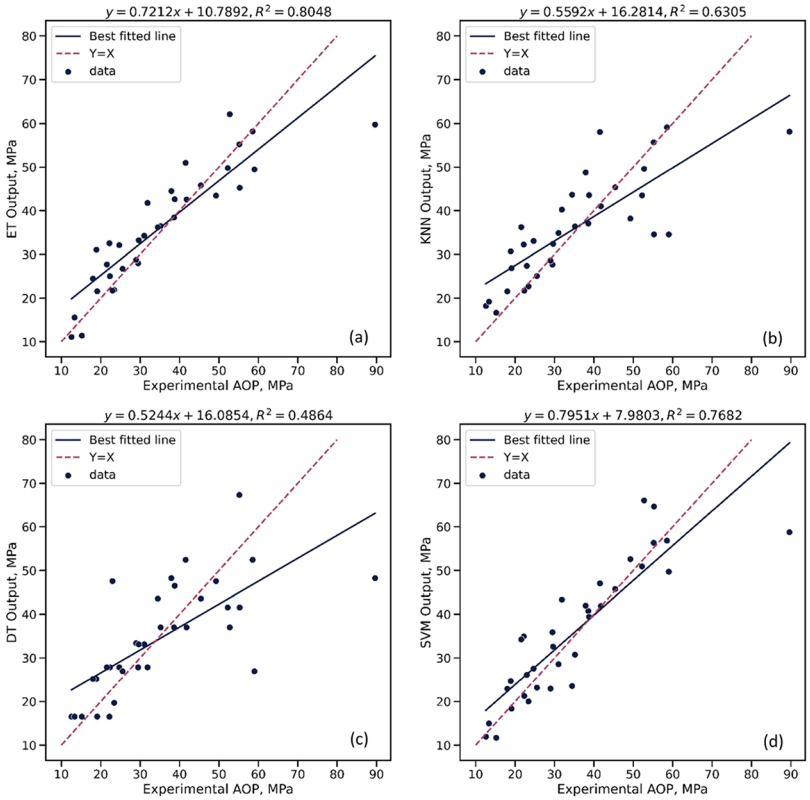



**Figure 3.** Predicted versus experimental values of AOP for (a) ET model, (b) KNN model, (c) DT model, and (d) SVM model in the testing data.



https://doi.org/10.1021/acsomega.2c03192
ACS Omega 2022, 7, 30113−30124

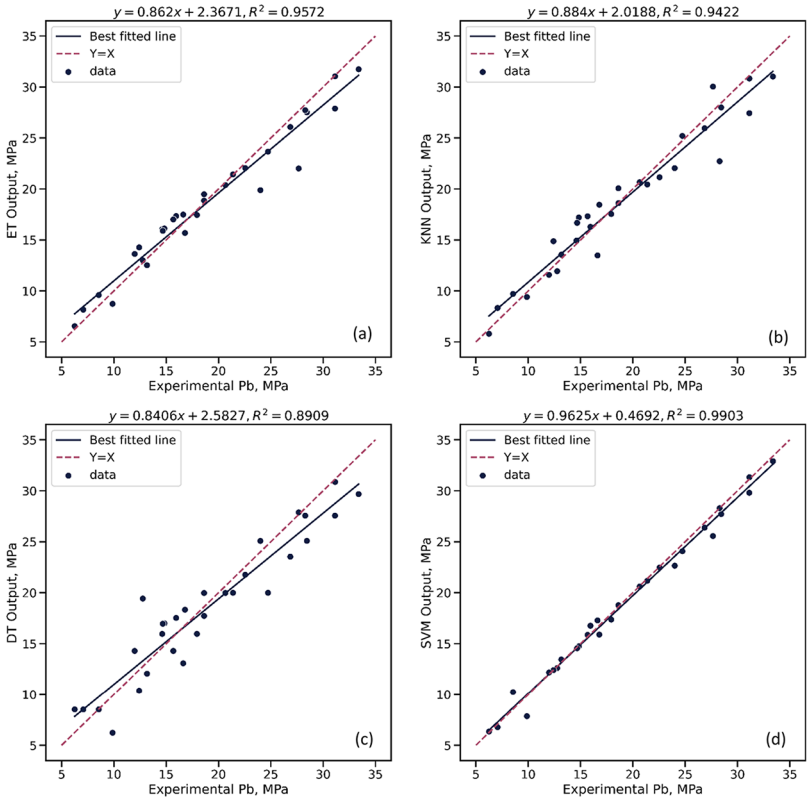



**Figure 4.** Predicted versus experimental values of Pb for (a) ET model, (b) KNN, (c) DT model, and (d) SVM model in the testing data.

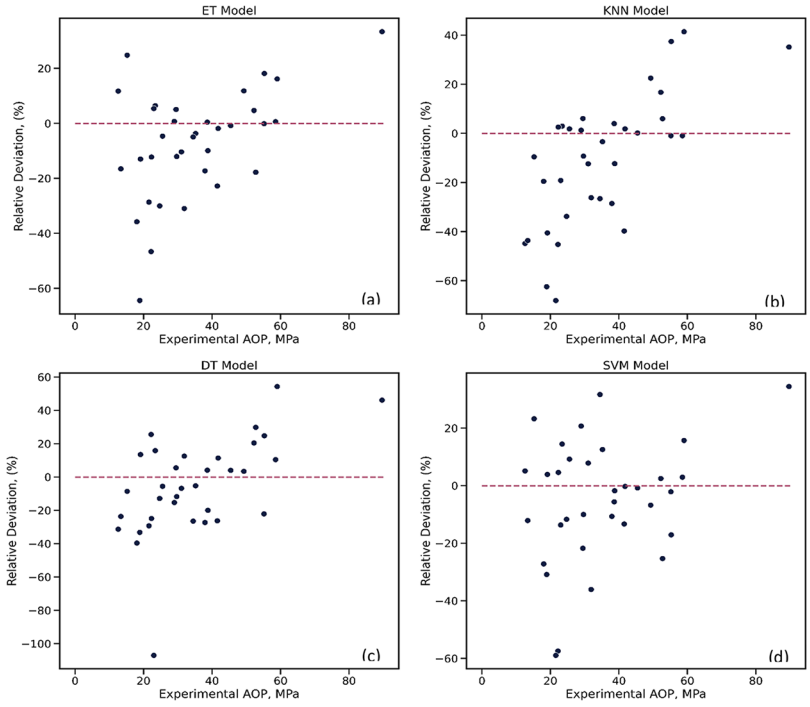



**Figure 5.** Relative deviation of the model predictions in the testing data set for AOP: (a) ET model, (b) KNN model, (c) DT model, and (d) SVM model.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (2)$$

where $\hat{y}_i$ and $\overline{y}$ are the predicted value for the $i$th sample and average values, respectively.

It should be emphasized here that we used only the testing data set to assess the performance of the models to avoid bias in estimating these indices that could mislead evaluation of the models. The training data set was used only for training the models.



https://doi.org/10.1021/acsomega.2c03192
ACS Omega 2022, 7, 30113−30124

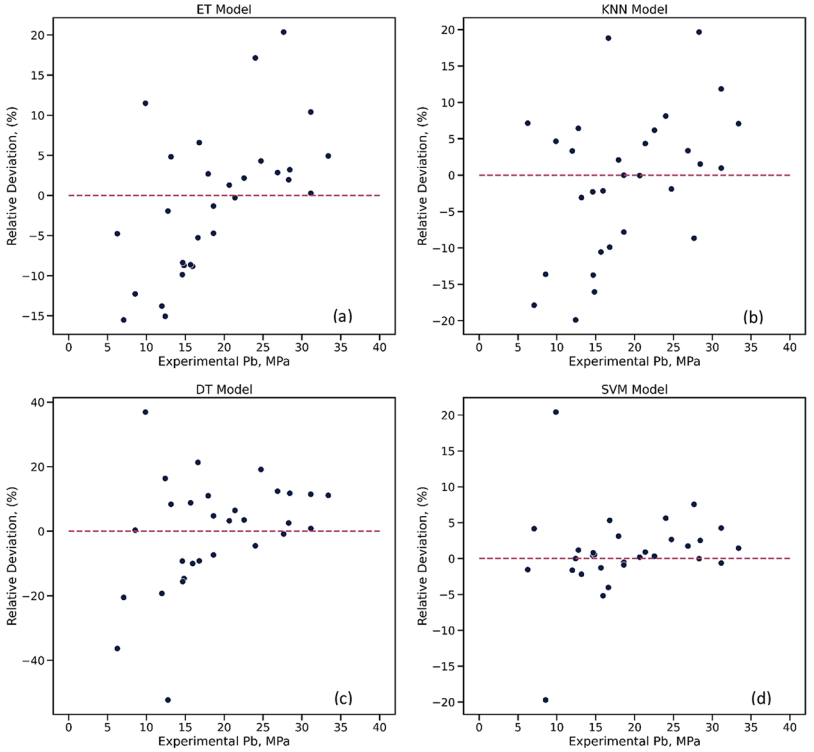



**Figure 6.** Relative deviation of the model predictions in the testing data set for Pb: (a) ET model, (b) KNN model, (c) DT model, and (d) SVM model.

The reliability of the developed models was further tested using some noisy data samples. Twenty percent of the total data (34 samples for the AOP data set and 29 samples for the Pb data set) was randomly selected. Then, a Gaussian noise (with mean equal to zero and specified standard deviation) was added to each feature in this data set for generating the noisy samples. The standard deviation of the noises was determined based on standard deviation of each feature (2% changes) in the randomly selected data samples for creating feasible values. Finally, the AI/ML model results (e.g., ET, DT, KNN, and SVM) were compared with the PC-SAFT model results to evaluate accuracy of the AI/ML models in this research work.

*2.2.9. PC-SAFT Model Characterization.* The SAFT model assumes that all molecules are spherical segments with equal sizes. This method considers five main pure components including segment number ($m$), segment diameter ($\sigma$), segment energy ($\varepsilon$), associate energy ($\varepsilon_{A_iB_i}$), and associate volume ($\kappa_{A_iB_i}$). Associate energy and associate volume are important when self-associating is considered between molecules. The SAFT model is modified by considering dipole−dipole interactions using the Lennard−Jones equation for calculation of the segment contribution. This modified model is called PC-SAFT.[54] The PC-SAFT model successfully can be used to model chain molecules like hydrocarbon solution because of considering dispersion interaction between them. Helmholtz free energy is used to calculate thermodynamic properties. Hence, the PC-SAFT model uses dimensionless residual Helmholtz energy to estimate thermodynamic properties. The following equation shows this energy with considering different types of interactions.[31,55]

$$\tilde{a}^{res} = \tilde{a}^{hc} + \tilde{a}^{disp} + \tilde{a}^{assoc} \tag{3}$$

where dispersion forces ($\tilde{a}^{disp}$) show the interaction among nonpolar and weakly polar molecules. The hydrogen interaction and donor-acceptor forces show using association contribution ($\tilde{a}^{assoc}$), while hard-chain formation and sphere contribution demonstrate using hard-chain contribution ($\tilde{a}^{hc}$). These parameters and residual Helmholtz energy ($\tilde{a}^{res}$) are calculated using eqs 45678 as follows:

$$\left(\frac{\tilde{a}^{hc}}{RT}\right) = \left(\frac{a_0^{hs}}{RT}\right)\sum_i X_i m_i + \frac{a^{assoc}}{RT} \tag{4}$$

$$\left(\frac{\tilde{a}^{assoc}}{RT}\right) = \sum_i X_i(1-m_i)\ln g_{ii}(d_{ii}) \tag{5}$$

$$\left(\frac{\tilde{a}_0^{hs}}{RT}\right) = \frac{1}{\zeta_0}\left[\frac{3\zeta_1\zeta_2}{(1-\zeta_3)} + \frac{\zeta_2^3}{\zeta_3(1-\zeta_3)^2} + \left(\frac{\zeta_2^3}{\zeta_3^2} - \zeta_0\right)\ln(1-\zeta_3)\right] \tag{6}$$

$$d_{ii} = \sigma_i\left[1 - 0.12\exp\left(-\frac{3\in_{ii}}{k_{BT}T}\right)\right] \tag{7}$$

$$\tilde{a}^{res} = \frac{A^{res}}{NkT} \tag{8}$$

where $A_{res}$ denotes the residual Helmholtz free energy. $N$, $k$, and $T$ are the number of molecules, Boltzmann constant, and absolute temperature, respectively. $d$, $x$, and $m_i$ are the sphere diameter, molar fraction of a chain component, and number of segments in the chain, respectively.



https://doi.org/10.1021/acsomega.2c03192
ACS Omega 2022, 7, 30113−30124

Table 6. Gaussian Noise Information for the AOP and Pb Data Sets

| feature | Gaussian noise standard deviation for the AOP data | Gaussian noise standard deviation for the Pb data |
|---|---|---|
| reservoir pressure | 0.219 | 0.229 |
| reservoir temperature | 0.371 | 0.380 |
| API gravity | 0.116 | 0.126 |
| GOR | 1.196 | 1.301 |
| fluid molecular weight | 0.472 | 0.415 |
| oil-$N_2$ | 0.027 | 0.005 |
| oil-$CO_2$ | 0.018 | 0.022 |
| oil-$H_2S$ | 0.025 | 0.028 |
| oil-$C_1$ | 0.202 | 0.225 |
| oil-$C_2$ | 0.042 | 0.042 |
| oil-$C_3$ | 0.024 | 0.027 |
| oil-i-$n$-$C_4$ | 0.043 | 0.045 |
| oil-i-$n$-$C_5$ | 0.030 | 0.030 |
| oil-$C_6$ | 0.031 | 0.030 |
| saturates | 0.252 | 0.252 |
| aromatics | 0.164 | 0.150 |
| resins | 0.077 | 0.080 |
| asphaltenes | 0.097 | 0.104 |
| $N_2$ | 0.047 | 0.002 |
| $CO_2$ | 0.201 | 0.125 |
| $H_2S$ | 0.001 | 0.037 |
| $C_1$ | 0.156 | 0.172 |
| $C_2$ | 0.031 | 0.022 |
| $C_3$ | 0.021 | 0.014 |
| $C_{4+}$ | 0.015 | 0.015 |
| temperature | 0.637 | 0.634 |

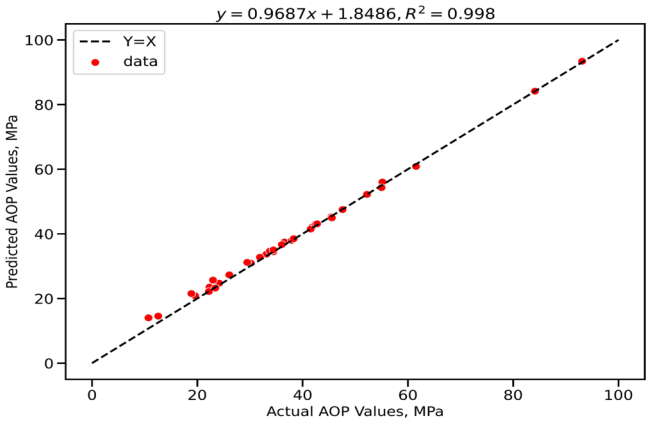



Figure 7. Performance of the extra trees model in prediction of the AOP values using the noisy data samples.

## 3. RESULTS AND DISCUSSION

In this section, the results of the developed models are presented and discussed. Then, the modeling results are compared with the experimental data and the PC-SAFT model.

**3.1. Data Statistics.** Statistical information for the features and the target variables for AOP and Pb is presented in Tables 2 and 3, respectively.

**3.2. Statistical Analysis.** Performances of the developed models in predicting AOP and Pb are summarized in Tables 4 and 5. The fitting and evaluation process was performed 100 times for each model to test repeatability of the results. Based on Tables 4 and 5, the ET ($R^2$ = 0.793) and the SVM ($R^2$ = 0.988) models are the best predictor of AOP and Pb, respectively. In general, the error analysis performance for the testing results in Tables 4 and 5 shows that the developed AI/ML methods are more reliable in predicting Pb compared to AOP.

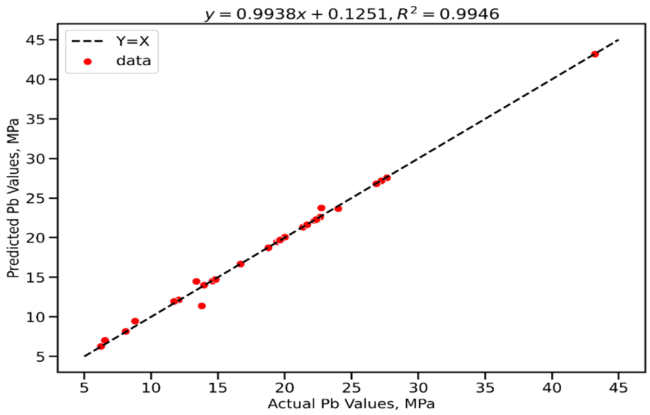



Figure 8. Accuracy of the support vector machine model in prediction of the Pb values using the noisy data samples.

**3.3. Graphical Analysis.** Graphical analysis was carried out to understand performance of the developed AI/ML models. A comparison between the predicted and experimental values of AOP and Pb in the testing phase is presented in Figures 3 and 4. Based on Figures 3 and 4, panels (a)−(d), the ET and the SVM models have the highest accuracy in the estimation of AOP and Pb, respectively.

The residual analysis that presents relative deviation (%) of the predicted AOP and Pb values and the experimental data for the testing data set are presented in Figures 5 and 6. Equation 9 was used to determine the relative deviation as follows:

$$\text{relative deviation (\%)} = \frac{y_{\text{exp}} - y_{\text{pred}}}{y_{\text{exp}}} \times 100 \quad (9)$$

where $y_{\text{exp}}$ and $y_{\text{pred}}$ are experimental and predicted values, respectively. Based on Figures 5 and 6, the ET model yielded the best performance and the majority of the deviations accumulated between −20% and 20%. In addition, the SVM model provided an accurate efficiency for Pb (deviation between −5% and 5%).

**3.4. Reliability Test.** According to the graphical and statistical analyses, the SVM and the ET models showed the best performance in estimating Pb and AOP, respectively. However, the model was tested using noisy data from the original data set. A random 20% of the original data set was chosen, and noisy samples were generated by adding Gaussian noise to each feature in the data set. The information related to the noises that are added to each feature is presented in Table 6.

Efficiency of the ET and the SVM models (the best models) in estimating AOP and Pb parameters, respectively, using the noisy data samples is presented in Figures 7 and 8. These figures confirm the accuracy of the developed models in prediction of the target values.

**3.5. Comparison between Smart Models and the PC-SAFT.** The results from comparing the experiment and AI/ML models in estimating the AOP and Pb with the results of the PC-SAFT model reported by Panuganti et al. are presented in Table 7.[15] Based on the results, the PC-SAFT model has a better performance in predicting AOP during gas injection


https://doi.org/10.1021/acsomega.2c03192
ACS Omega 2022, 7, 30113−30124

Table 7. Compression between the Experimental Data, the PC-SAFT Model, and the AI/ML Models Developed in This Research Work

| | AOP | | | | | | Pb | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | experimental | PC-SAFT | ET | KNN | DT | SVM | experimental | PC-SAFT | ET | KNN | DT | SVM |
| 5% gas injection | 3414.94 | 3414.94 | 2473.64 | 1925.92 | 1806.88 | 2723.20 | 1761.71 | 2003.65 | 1520.96 | 1559.37 | 1483.48 | 1696.12 |
| | 2766.24 | 2604.95 | 2005.80 | 1563.16 | 1466.70 | 2207.42 | 1959.79 | 2121.08 | 1691.71 | 1734.74 | 1649.98 | 1886.77 |
| | 3397.26 | 3360.37 | 2460.89 | 1916.03 | 1797.61 | 2709.14 | 2155.25 | 2301.37 | 1860.19 | 1907.26 | 1814.29 | 2074.90 |
| | 2958.90 | 2776.26 | 2144.75 | 1670.90 | 1567.73 | 2360.60 | 2301.37 | 2447.49 | 1986.15 | 2036.43 | 1937.11 | 2215.54 |
| 15% gas injection | 4285.71 | 4489.80 | 3101.64 | 2412.85 | 2263.51 | 3415.55 | 2285.71 | 2489.80 | 1972.65 | 2022.59 | 1923.95 | 2200.47 |
| | 3714.29 | 4040.82 | 2689.54 | 2093.31 | 1963.86 | 2961.21 | 2775.51 | 2693.88 | 2395.86 | 2455.57 | 2335.68 | 2671.90 |
| | 4740.74 | 4740.74 | 3429.81 | 2667.3 | 2502.13 | 3777.34 | 2444.44 | 2481.48 | 2109.47 | 2162.90 | 2057.38 | 2353.24 |
| | 4037.04 | 3851.85 | 2922.30 | 2273.79 | 2133.11 | 3217.83 | 2518.52 | 2592.59 | 2173.33 | 2228.39 | 2119.65 | 2424.54 |
| | 3703.70 | 3592.59 | 2681.90 | 2087.39 | 1958.31 | 2952.79 | 2888.89 | 2814.81 | 2492.59 | 2555.80 | 2430.98 | 2781.03 |
| 30% gas injection | 9183.67 | 9183.67 | 6634.05 | 5151.79 | 4832.00 | 7309.92 | 3020.41 | 3020.41 | 2605.96 | 2672.06 | 2541.54 | 2907.61 |
| | 7020.41 | 6489.80 | 5073.91 | 3942.09 | 3697.59 | 5589.91 | 3428.57 | 3265.31 | 2957.79 | 3032.87 | 288.64 | 3300.47 |
| | 8444.44 | 8222.22 | 6100.92 | 4738.41 | 4444.35 | 6722.15 | 3037.04 | 3296.30 | 2620.30 | 2686.76 | 2555.52 | 2923.62 |
| | 6185.19 | 6370.37 | 4471.55 | 3475.04 | 3259.60 | 4925.82 | 3555.56 | 3370.37 | 3067.26 | 3145.13 | 2991.39 | 3422.70 |
| $R^2$ | | 0.99 | 0.79 | 0.62 | 0.48 | 0.76 | | 0.99 | 0.97 | 0.94 | 0.88 | 0.99 |
| RMSE | | 1.50 | 7.51 | 10.16 | 11.87 | 7.99 | | 1.07 | 1.71 | 1.84 | 2.51 | 0.79 |

compared to AI/ML models. As mentioned in the previous section, accuracy of the ET ($R^2$: 0.79) is higher than that of the SVM ($R^2$: 0.76), the KNN ($R^2$: 0.62), and the DT ($R^2$: 0.48) for AOP. Hence, more input data and experimental points should be used in the AI/ML models for prediction of the AOP to minimize the deviation from actual data. Based on Table 7, the results of the AI/ML models in prediction of Pb were acceptable compared to the PC-SAFT model. The PC-SAFT model ($R^2$: 0.99) has similar accuracy to the SVM ($R^2$: 0.99) model in prediction of Pb. In addition, the results of the ET ($R^2$: 0.97) and KNN ($R^2$: 0.94) are reliable for Pb prediction. The amounts of $R^2$ and RMSE for the AI/ML models presented in Table 7 are based on all data points, not only data from Panuganti et al.[15] Hence, it can be concluded that the SVM, the ET, and the KNN models are a good alternative for the PC-SAFT model for prediction of Pb without using complex mathematical relations. However, in the case of AOP, the results of the PC-SAFT model are better than those of the developed AI/ML models. It is worth mentioning here that in the AI/ML models, data from different studies reported in the literature were used, while in the PC-SAFT model only data from Panuganti et al.[15] were considered. Hence, the accuracy of the PC-SAFT is higher than that of AI/ML models and has a lower probable error. As a result, for prediction of Pb, the AI/ML model is a reliable alternative to the PC-SAFT model and can predict the results successfully without considering complex equations and computational efforts. In addition, the AI/ML model can calculate parameters in various ranges of temperature, while the PC-SAFT model has a poor performance at low temperatures.[9]

Although numerous experimental and theoretical research studies attempted to predict the pressure that asphaltene starts to precipitate during gas injection, there are still several challenges associated with these techniques. Some of the challenges include high computational requirements and efforts, high cost, and the need to solve complex mathematical relationships. It can be concluded here that soft computing methods can assist in overcoming the limitations and estimating target values in a simpler way and with an acceptable error. However, the accuracy of smart methods highly depends on input parameters, which is the main drawback of these methods.

## 4. CONCLUSIONS

Pressure decline in oil reservoirs during gas injection is a problematic phenomenon that causes asphaltene precipitation and deposition leading to flow assurance issues. Reservoir temperature, reservoir pressure, SARA fraction, API gravity, GOR, fluid molecular weight, monophasic composition, and composition of gas injection were determined by various researchers to be the parameters affecting estimation of the AOP and Pb. Hence, in this research work we attempted to develop four predictive AI/ML-based intelligent models (e.g., ET, DT, SVM, and KNN) to estimate AOP and Pb which are considered as critical parameters in determining the onset of asphaltene precipitation in oil reservoirs. A comprehensive database was collected from experimental research studies reported in the literature for this aim (AOP: 170 data points and Pb: 146 data points). The data set used in the present research work is relatively large and covers all available data from different sources. Missing value imputation, data statistics, and data scaling are three main preparing data methods that we used for preprocessing the database for AI/ML modeling. The following main five conclusions can be drawn from this research work:

1- The error analysis and graphical studies showed that the ET ($R^2$: 0.79, RMSE: 7.5) and the SVM ($R^2$: 0.98, RMSE: 0.76) are the best predictors in estimating AOP and Pb, respectively. In addition, the results of the DT method showed notable deviation from the real data for both data sets. The accuracy of the developed smart models for prediction of the AOP and Pb in the present research work is in the following order:
   AOP: ET > SVM > KNN > DT
   Pb: SVM > ET > KNN > DT
2- The reliability test results using randomly generated noisy data confirmed accuracy of the developed ET ($R^2$: 0.998) and the SVM ($R^2$: 0.994) models in predicting the AOP and Pb, respectively, compared to the AI/ML models used in the present research work.
3- Generally, the accuracy of the PC-SAFT model is more than that of the AI/ML models. However, our results confirm that the AI/ML approach is an acceptable alternative for the PC-SAFT model when we face lack of data and/or complex mathematical equations.


https://doi.org/10.1021/acsomega.2c03192
ACS Omega 2022, 7, 30113−30124

4- The error analysis results revealed that the AI/ML model performance in predicting the Pb is more reliable than the AOP.
5- The outcomes of the present research work show a promising pathway in estimating the probability of asphaltene precipitation during gas injection using smart computational techniques. The gas injection rate can affect the AOP and Pb. Hence, it would be interesting to consider this parameter in future AI/ML modeling efforts.

## 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

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

**Abdollah Davoudi** − *Department of Petroleum Engineering, School of Chemical and Petroleum Engineering, Shiraz University, Shiraz 71348-14336, Iran*

**Hossein Parsaei** − *Department of Medical Physics and Engineering, School of Medicine, Shiraz University of Medical Sciences, Shiraz 71348-14336, Iran*

**Timur Sh. Atabaev** − *Department of Chemistry, Nazarbayev University, Nur-Sultan 010000, Kazakhstan*

**Oleksandr P. Ivakhnenko** − *Department of Petroleum Engineering, Kazakh British Technical University, Almaty 050000, Kazakhstan*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c03192


### Author Contributions

S.T.: Investigation, methodology, formal analysis, visualization, validation, and writing—original draft preparation. A.D.: Investigation, methodology, formal analysis, visualization, validation, and writing—original draft preparation. A.S.: Conceptualization, supervision, methodology, formal analysis, funding acquisition, project administration, resources, data curation, and writing—review and editing. H.P.: Methodology, formal analysis, visualization, validation, writing—original draft preparation, review, and editing. T.S.A.: Methodology, formal analysis, funding acquisition, and writing—review and editing. O.P.I.: Methodology, formal analysis, funding acquisition, and writing—review and editing.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

A.S., T.Sh.A., and O.P.I. acknowledge the financial support received for their research from Nazarbayev University through a Collaborative Research Proposal (grant# 091019CRP2103). This CRP grant made possible official collaboration between Nazarbayev University and Kazakh British Technical University. The authors wish to thank the editors and two anonymous reviewers for their critical yet fair and constructive comments, which helped the authors to improve the quality and clarity of the manuscript. The authors also would like to thank Asset Abenov and Nurzhan Seitmaganbetov for their assistance in data extraction.

## NOMENCLATURE

### Acronyms

| | |
|---|---|
| ANN | artificial neural network |
| AOP | asphaltene onset pressure |
| CPA | Cubic Plus Association |
| DT | decision tree |
| EOS | equation of state |
| GA | genetic algorithm |
| GOR | gas−oil ratio |
| LAOP | lower asphaltene onset pressure |
| LSSVM | least-squares support vector machine |
| MELM | multilayer extreme learning machine |
| OBM | oil-based mud |
| OFVF | oil formation volume factor |
| Pb | bubble point pressure |
| PC-SAFT | Perturbed-chain form statistical associating fluid theory |
| PR | Peng Robinson |
| PSO | particle swarm optimization |
| QCR | quartz crystal resonator |
| RF | random forest |
| SAFT | Statistical Associating Fluid Theory |
| SRK | Soave−Redlich−Kwong |
| STO | stock tank oil |
| SVM | support vector machine |
| SVR | support vector regression |
| UAOP | upper asphaltene onset pressure |

### Variables and Parameters

| | |
|---|---|
| $\tilde{a}^{res}$ | total residual Helmholtz free energy |
| $\tilde{a}^{disp}$ | force between nonpolar and weakly polar molecules |
| $\tilde{a}^{assoc}$ | association contribution |
| $\tilde{a}^{hc}$ | hard-chain contribution |
| $d$ | sphere diameter |
| $k$ | Boltzmann constant |
| $m_i$ | number of the segment in the chain |
| $N$ | number of molecules |
| $T$ | temperature |
| $x$ | molar fraction of chain component |

### Greek Letters

| | |
|---|---|
| $\varepsilon$ | segment energy |
| $\varepsilon_{A_iB_i}$ | associate energy |
| $\kappa_{A_iB_i}$ | associate volume |
| $\sigma$ | segment diameter |

## REFERENCES


(1) Rahmani, N. H. G.; Masliyah, J. H.; Dabros, T. Characterization of asphaltenes aggregation and fragmentation in a shear field. *AIChE J.* **2003**, *49*, 1645−1655.

(2) Nascimento, F. P.; Pereira, V. J.; Souza, R. P.; Lima, R. C. A.; Costa, G. M. N.; Rosa, P. T. V.; Forca, A. F.; Vieira de Melo, S. A. B. An experimental and theoretical investigation of asphaltene precipitation in a crude oil from the Brazilian pre-salt layer under $CO_2$ injection. *Fuel* **2021**, *284*, No. 118968.

(3) Gonzalez, D. L.; Vargas, F. M.; Hirasaki, G. J.; Chapman, W. G. Modeling Study of $CO_2$-Induced Asphaltene Precipitation. *Energy Fuels* **2008**, *22*, 757−762.

(4) Yonebayashi, H.; Miyagawa, Y.; Ikarashi, M.; Watanabe, T.; Maeda, H.; Yazawa, N. Determination of Asphaltene-Onset Pressure Using Multiple Techniques in Parallel. *SPE Prod. Oper.* **2018**, *33*, 486−497.

(5) Cardoso, F. M. R.; Carrier, H.; Daridon, J. L.; Pauly, J.; Rosa, P. T. V. $CO_2$ and Temperature Effects on the Asphaltene Phase Envelope As Determined by a Quartz Crystal Resonator. *Energy Fuels* **2014**, *28*, 6780−6787.
(6) Akbarzadeh, K.; Hammami, A.; Kharrat, A.; Zhang, D.; Allenson, S.; Creek, J.; Kabir, S.; Jamaluddin, A.; Marshall, A. G.; Rodgers, R. P.; Mullins, O. C.; Solbakken, T. Asphaltenes − problematic but rich in potential. *Oilfield Rev.* **2007**, *19*, 22−43.
(7) Alimohammadi, S.; Zendehboudi, S.; James, L. A comprehensive review of asphaltene deposition in petroleum reservoirs: Theory, challenges, and tips. *Fuel* **2019**, *252*, 753−791.
(8) Punnapala, S.; Vargas, F. M. Revisiting the PC-SAFT characterization procedure for an improved asphaltene precipitation prediction. *Fuel* **2013**, *108*, 417−429.
(9) Seitmaganbetov, N.; Rezaei, N.; Shafiei, A. Characterization of crude oils and asphaltenes using the PC-SAFT EoS: A systematic review. *Fuel* **2021**, *291*, No. 120180.
(10) Flory, P. J. Thermodynamics of High Polymer Solutions. *J. Chem. Phys.* **1941**, *9*, 660.
(11) Scatchard, G. Equilibria in Non-electrolyte Solutions in Relation to the Vapor Pressures and Densities of the Components. *Chem* **1931**, *8*, 321−333.
(12) Scott, R. L.; Magat, M. The Thermodynamics of High-Polymer Solutions: I. The Free Energy of Mixing of Solvents and Polymers of Heterogeneous Distribution. *J. Chem. Phys.* **1945**, *13*, 172−177.
(13) Ting, P. D.; Hirasaki, G. J.; Chapman, W. G. *Pet. Sci. Technol.* **2003**, 647−661.
(14) Gross, J.; Sadowski, G. Perturbed-Chain SAFT: An Equation of State Based on a Perturbation Theory for Chain Molecules. *Ind. Eng. Chem. Res.* **2001**, *40*, 1244−1260.
(15) Panuganti, S. R.; Vargas, F. M.; Gonzalez, D. L.; Kurup, A. S.; Chapman, W. G. PC-SAFT characterization of crude oils and modeling of asphaltene phase behavior. *Fuel* **2012**, *93*, 658−669.
(16) Gonzalez, D. L.; Ting, P. D.; Hirasaki, G. J.; Chapman, W. G. Prediction of Asphaltene Instability under Gas Injection with the PC-SAFT Equation of State. *Energy Fuels* **2005**, *19*, 1230−1234.
(17) Gonzalez, D. L.; Hirasaki, G. J.; Creek, J.; Chapman, W. G. Modeling of Asphaltene Precipitation Due to Changes in Composition Using the Perturbed Chain Statistical Associating Fluid Theory Equation of State. *Energy Fuels* **2007**, *21*, 1231−1242.
(18) Zendehboudi, S.; Shafiei, A.; Bahadori, A.; James, L. A.; Elkamel, A.; Lohi, A. Asphaltene precipitation and deposition in oil reservoirs − Technical aspects, experimental and hybrid neural network predictive tools. *Chem. Eng. Res. Des.* **2014**, *92*, 857−875.
(19) Khamehchi, E.; Behvandi, R.; rashidi, f. Prediction of Bubble Point Pressure & Asphaltene Onset Pressure During $CO_2$ Injection Using ANN & ANFIS Models. *J. Pet. Sci. Technol.* **2011**, *1*, 35−45.
(20) Kamari, A.; Safiri, A.; Mohammadi, A. H. Compositional Model for Estimating Asphaltene Precipitation Conditions in Live Reservoir Oil Systems. *J. Dispersion Sci. Technol.* **2015**, *36*, 301−309.
(21) Rashidi, S.; Mehrad, M.; Ghorbani, H.; Wood, D. A.; Mohamadian, N.; Moghadasi, J.; Davoodi, S. Determination of bubble point pressure & oil formation volume factor of crude oils applying multiple hidden layers extreme learning machine algorithms. *J. Pet. Sci. Eng.* **2021**, *202*, No. 108425.
(22) Jamaluddin, A. K. M.; Creek, J.; Kabir, C. S.; McFadden, J. D.; D'Cruz, D.; Joseph, M. T.; Joshi, N.; Ross, B. A Comparison of Various Laboratory Techniques to Measure Thermodynamic Asphaltene Instability. In *SPE Asia Pacific Improved Oil Recovery Conference*, 2001.
(23) Gonzalez, D. L.; Garcia, M. E.; Diaz, O. Unusual Asphaltene Phase Behavior of Fluids from Lake Maracaibo, Venezuela. In *SPE Latin America and Caribbean Petroleum Engineering Conference*, 2012.
(24) Ebrahimi, M.; Mousavi-Dehghani, S. A.; Dabir, B.; Shahrabadi, A. The effect of aromatic solvents on the onset and amount of asphaltene precipitation at reservoir conditions: Experimental and modeling studies. *J. Mol. Liq.* **2016**, *223*, 119−127.
(25) Fouad, W. A.; Abutaqiya, M. I. L.; Mogensen, K.; Yap, Y. F.; Goharzadeh, A.; Vargas, F. M.; Vega, L. F. Predictive Model for Pressure−Volume−Temperature Properties and Asphaltene Instability of Crude Oils under Gas Injection. *Energy Fuels* **2018**, *32*, 8318−8328.
(26) Negahban, S.; Bahamaish, J. N. M.; Joshi, N.; Nighswander, J.; Jamaluddin, A. K. M. An Experimental Study at an Abu Dhabi Reservoir of Asphaltene Precipitation Caused By Gas Injection. *SPE Prod. Facil.* **2005**, *20*, 115−125.
(27) Abutaqiya, M. I. L.; Sisco, C. J.; Khemka, Y.; Safa, M. A.; Ghloum, E. F.; Rashed, A. M.; Gharbi, R.; Santhanagopalan, S.; Al-Qahtani, M.; Al-Kandari, E.; Vargas, F. M. Accurate Modeling of Asphaltene Onset Pressure in Crude Oils Under Gas Injection Using Peng−Robinson Equation of State. *Energy Fuels* **2020**, *34*, 4055−4070.
(28) Vargas, F. M.; Garcia-Bermudes, M.; Boggara, M.; Punnapala, S.; Abutaqiya, M.; Mathew, N.; Prasad, S.; Khaleel, A.; Al Rashed, M.; Al Asafen, H. On the Development of an Enhanced Method to Predict Asphaltene Precipitation. In *Offshore Technology Conference*, 2014.
(29) Abutaqiya, M. I. L.; Sisco, C. J.; Wang, J.; Vargas, F. M. Systematic Investigation of Asphaltene Deposition in the Wellbore and Near-Wellbore Region of a Deepwater Oil Reservoir Under Gas Injection. Part 1: Thermodynamic Modeling of the Phase Behavior of Polydisperse Asphaltenes. *Energy Fuels* **2019**, *33*, 3632−3644.
(30) Buenrostro-Gonzalez, E.; Lira-Galeana, C.; Gil-Villegas, A.; Wu, J. Asphaltene precipitation in crude oils: Theory and experiments. *AIChE J.* **2004**, *50*, 2552−2570.
(31) Masoudi, M.; Miri, R.; Hellevang, H.; Kord, S. Modified PC-SAFT characterization technique for modeling asphaltenic crude oil phase behavior. *Fluid Phase Equilib.* **2020**, *513*, No. 112545.
(32) Gonzalez, D. L.; Mahmoodaghdam, E.; Lim, F.; Joshi, N. Effects of Gas Additions to Deepwater Gulf of Mexico Reservoir Oil: Experimental Investigation of Asphaltene Precipitation and Deposition. In *SPE Annual Technical Conference and Exhibition*, 2012.
(33) Cañas-Marín, W. A.; González, D. L.; Hoyos, B. A. A theoretically modified PC-SAFT equation of state for predicting asphaltene onset pressures at low temperatures. *Fluid Phase Equilib.* **2019**, *495*, 1−11.
(34) Bahrami, P.; Kharrat, R.; Mahdavi, S.; Firoozinia, H. Prediction of the gas injection effect on the asphaltene phase envelope. *Oil Gas Sci. Technol. − Rev. IFP Energies nouvelles* **2015**, *70*, 1075−1086.
(35) Memon, A.; Qassim, B.; Al-Ajmi, M.; Kumar Tharanivasan, A.; Gao, J.; Ratulowski, J.; Al-Otaibi, B.; Khan, R. A. Miscible Gas Injection and Asphaltene Flow Assurance Fluid Characterization: A Laboratory Case Study for a Black Oil Reservoir. In *SPE EOR Conference at Oil and Gas West Asia*, 2012.
(36) Jamaluddin, A. K. M.; Nighswander, J. N.; Kohse, B. F.; El Mahdi, A.; Binbrek, M. A.; Hogg, P. F. Experimental and Theoretical Assessment of the Asphaltene Precipitation Characteristics of the Sahil Field Under a Proposed Miscible Gas Injection Scheme. In *Abu Dhabi International Petroleum Exhibition and Conference*, 2000.
(37) Jamaluddin, A. K. M.; Nighswander, J.; Joshi, N. A Systematic Approach for Characterizing Hydrocarbon Solids. *SPE J.* **2003**, *8*, 304−312.
(38) Sullivan, M.; Smythe, E. J.; Fukagawa, S.; Harrison, C.; Dumont, H.; Borman, C. A Fast Measurement of Asphaltene Onset Pressure. *SPE Reservoir Eval. Eng.* **2020**, *23*, 0962−0978.
(39) Abdallah, D. Impact of Asphaltenes Deposition on Completion Design for $CO_2$ Pilot in an Onshore Abu Dhabi Field. In *Abu Dhabi International Petroleum Conference and Exhibition*, 2012.
(40) Jamaluddin, A. K. M.; Joshi, N.; Iwere, F.; Gurpinar, O. An Investigation of Asphaltene Instability Under Nitrogen Injection. In *SPE International Petroleum Conference and Exhibition in Mexico*, 2002.
(41) 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.
(42) Kokal, S. L.; Najman, J.; Sayegh, S. G.; George, A. E. Measurement And Correlation Of Asphaltene Precipitation From


https://doi.org/10.1021/acsomega.2c03192
*ACS Omega* 2022, 7, 30113−30124

Heavy Oils By Gas Injection. *J. Can. Pet. Technol.* **1992**, *31*, DOI: 10.2118/92-04-01.
(43) Hu, Y.-F.; Li, S.; Liu, N.; Chu, Y.-P.; Park, S. J.; Mansoori, G. A.; Guo, T.-M. Measurement and corresponding states modeling of asphaltene precipitation in Jilin reservoir oils. *J. Pet. Sci. Eng.* **2004**, *41*, 169−182.
(44) 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.
(45) 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.
(46) Furey, T. S.; Cristianini, N.; Duffy, N.; Bednarski, D. W.; Schummer, M.; Haussler, D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* **2000**, *16*, 906−914.
(47) Imandoust, S. B.; Bolandraftar, M. Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *Int. J. Eng. Res.* **2013**, *3*, 605−610.
(48) Sattari, M.; Kamari, A.; Mohammadi, A. H.; Ramjugernath, D. Rigrous model development for estimating asphaltene precipitation from crude oil by n-alkane titration. *Pet. Coal* **2019**, *61*, 998−1008.
(49) Amiri-Ramsheh, B.; Safaei-Farouji, M.; Larestani, A.; Zabihi, R.; Hemmati-Sarapardeh, A. Modeling of wax disappearance temperature (WDT) using soft computing approaches: Tree-based models and hybrid models. *J. Pet. Sci. Eng.* **2022**, *208*, No. 109774.
(50) 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.
(51) Wu, J.; Chen, X.-Y.; Zhang, H.; Xiong, L.-D.; Lei, H.; Deng, S.-H. Hyperparameter Optimization for Machine Learning Models Based on Bayesian Optimizationb. *J. Electron. Sci. Technol.* **2019**, *17*, 26−40.
(52) Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M., Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Association for Computing Machinery: Anchorage, AK, USA, 2019; pp 2623−2631.
(53) 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.
(54) Huang, S. H.; Radosz, M. Equation of state for small, large, polydisperse, and associating molecules. *Ind. Eng. Chem. Res.* **1990**, *29*, 2284−2294.
(55) Kondori, J.; Zendehboudi, S.; James, L. Evaluation of Gas Hydrate Formation Temperature for Gas/Water/Salt/Alcohol Systems: Utilization of Extended UNIQUAC Model and PC-SAFT Equation of State. *Ind. Eng. Chem. Res.* **2018**, *57*, 13833−13855.


https://doi.org/10.1021/acsomega.2c03192
*ACS Omega* 2022, 7, 30113−30124