Article

Prediction of Foam Rheology Models Parameters Utilizing Machine Learning Tools

Jawad Al-Darweesh, Murtada Saleh Aljawad,* Zeeshan Tariq,* Shabeeb Alajmei, Bicheng Yan, and Muhammad Shahzad Kamal

Cite This: ACS Omega 2024, 9, 20397−20409

Read Online

ACCESS | 📊 Metrics & More | 📖 Article Recommendations

ABSTRACT: Rheological models are usually used to predict foamed fluid viscosity; however, obtaining the model constants under various conditions is challenging. Hence, this paper investigated the effect of different variables on foam rheology, such as shear rate, temperature, pressure, surfactant types, gas phase, and salinity, using a high-pressure high-temperature foam rheometer. Power-law, Bingham plastic, and Casson fluid models fit the experimental data well. Therefore, the data were fed to different machine learning techniques to evaluate the rheological model constants with different features. In this study, seven different machine learning techniques have been applied to predict the rheological models' constants, including decision tree, random forest, XGBoost (XGB), adaptive gradient boosting, gradient boosting, support vector regression, and voting regression. We evaluated the performance of our machine learning models using the coefficient of determination (R^2), cross-plots, root-mean-square error, and average absolute percentage error. Based on the prediction outcomes, the XGB model outperformed the other ML models. The XGB model exhibited remarkably low error rates, achieving a prediction accuracy of 95% under ideal conditions. Furthermore, our prediction results demonstrated that the Casson model accurately captured the rheological behavior of the foam. Additionally, we used Pearson's correlation coefficients to assess the significance of various properties in relation to the constants within the rheological models. It is evident that the XGB model makes predictions with nearly all features contributing significantly, while other machine learning techniques rely more heavily on specific features over others. The proposed methodology can minimize the experimental cost of measuring rheological parameters and serves as a quick assessment tool.
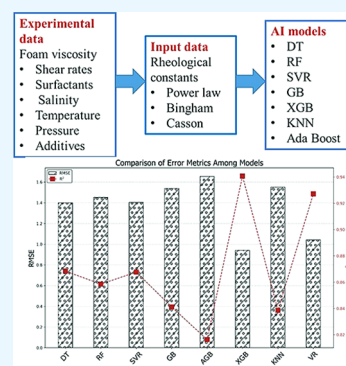
1. INTRODUCTION

The injection of water-based acid fracturing fluids in depleted carbonate reservoirs might hinder the recovery of treatment and reservoir fluids.[1] The flowback process is challenging because of the trapping of large liquid volumes into the formation due to the higher interfacial tension and low reservoir energy.[2] To overcome this challenge, foamed acid was introduced to energize the well and ease the flowback process. Additionally, foamed acid has successfully delivered deeper penetration, controlled the reaction, reduced acid loss, and improved acid diversion.[3] This was possible due to the creation of a multiphase fluid system where the gas is trapped between liquid lamella.

The success of the acid fracturing process strongly depends on the fluid's viscosity. It controls the fluid loss, reaction rate, and etching pattern.[4] Therefore, accurately determining acid foam viscosity is critical for effective acid fracturing techniques.[5] Foaming acid is classified as a non-Newtonian fluid due to its response to shearing. The rheological models, such as power law, Bingham plastic, or Herschel−Bulkley, are determined using the relationship between shear rate and stress or viscosity.[6] Generally, most of the published reports

show that foam fluid follows a power-law model.[6−9] However, few studies presented foam as Bingham plastic, Herschel−Bulkley, or Cross models.[10,11]

Einstein[1,15] derived the first model to determine the rheological behavior of foam. However, it is limited only to low foam qualities, below 10%. The minimum foam quality is 52%; below this range, it is known as an energized fluid. Mooney[2,16] developed an exponential model to measure the relative viscosity at high foam quality. The model was capable of measuring the viscosity within an acceptable range. However, it is limited to spherical bubbles. Similarly, Frankel and Acrivos[3,17] derived a suspension model that can only predict the foam viscosity for uniform spherical bubbles. Princen and Kiss[4,18] derived their model, considering the interaction between bubbles, bubble shape alteration, and
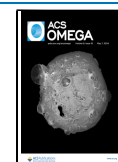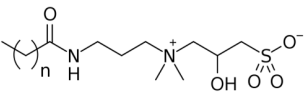
Table 1. Chemistry Details of Selected Surfactants

Surfactant	Molecular Structure	Charge	Manufacturer
Duomeen TTM	$R-N-CH_2-CH_2-CH_2-N-CH_3$ (with CH_3 groups on each N)	Cationic at low pH	Nouryon
Ethoduomeen T/13	$R-N-CH_2-CH_2-CH_2-N-CH_3$ (with CH_3CH_2OH groups on each N)	Cationic at low pH	Nouryon
Ethomeen C/12	$R-N-CH_3CH_2OH$ (with CH_3CH_2OH group on N)	Cationic at low pH	Nouryon
Armogel O	$CH_3-O-R-NH_2$	Cationic at low pH	Nouryon
Armovis VES	$R^1-C(=O)-NH(CH_2)_kN^+(CH_2)_mCH(CH_2)_nSO_3^-$ (with R^2, R^4 above and R^3 below)	Amphoteric	AkzoNobel
CAS 50	(structure drawing)	Amphoteric	CRODA

shear-thinning behavior. However, foam quality is not the only parameter controlling foam viscosity. The performance of foam viscosity is strongly related to the bubble size, bubble density, and bubble shapes and its response to the deformation in the dynamic state.[19] Additionally, the experimental studies showed that foam viscosity varied with many parameters, including foam quality, pressure, temperature, shear rate, salinity, and additives.[20−24]

Foam quality is the major factor affecting the performance of foam viscosity. It is described as the proportion of the gas volume fraction to the entire volume fraction (gas and liquid). Low foam quality results in low viscosity due to the low interaction among bubbles. While at high foam quality, the bubble number increases with significant interaction, which results in higher foam viscosity. In addition, the shear rate has the main effect on the foam viscosity. Following the power-law model, Al-Darweesh et al.[19] found that the foam viscosity decreased exponentially with the increase in shear rates. This observation was due to the deformation and shrinkage of bubbles during the shearing and, consequently, the viscosity reduction of the generated foam. Literature review[19,25] indicated that the foam viscosity drops remarkably at high temperatures due to the thermal thinning of foam film (lamella). This caused faster coarsening and rupture of bubbles.

Several surfactants generated foam at harsh reservoir conditions; however, most chemicals are degraded. Coconut monoethanolamide and Lauryl diethanolamide were not able to generate foam at a temperature of 60 °C.[25] Additionally, internal olefin sulfonate and sodium dodecyl sulfonate produced weak foam with salinity.[25,26] Due to their significant effects, additives such as corrosion inhibitors (CIs) are excessively utilized in well-stimulation techniques. However, CIs usually decrease the viscosity of viscoelastic surfactant solutions due to the breakage of large micelles into smaller ones.[27] In addition, Al-Darweesh et al.[20] reported the impact of CI on the half-life and viscosity of foam at high pressure and temperature. The results showed that the stability of the foam was reduced by 74% once CI was added to the Armovis solution. In addition, it reduced the viscosity by 40% at 500 1/

s. Carbon dioxide (CO_2) and nitrogen (N_2) are the most commonly used to generate foam. CO_2 is less stable than N_2 foam due to its high solubility within the liquid film.[28,29] However, CO_2 foam is more suitable for hot reservoirs since it is compatible with many fracturing fluids. CO_2 can interface with the cross-linking fluid structure by reducing the fluid pH.[2] Therefore, these parameters should be considered to predict accurate foam viscosity.

Due to the complex nature and poor understanding of foam behavior, the prediction of foam viscosity under harsh reservoir conditions is still a challenging task.[9,12−14] Recently, artificial intelligence (AI) tools have attracted special interest in oil and gas applications due to their ability to solve nonlinear problems with high accuracy. Othman et al.[30] predicted the fracturing fluid viscosity by considering different parameters such as polymer, temperature, and salinity, utilizing 86 experiments. They used six machine learning models, including decision trees (DTs), fully connected neural networks (FCNN), gradient boosting (GB), random forest (RF), extreme GB (XGB), and adaptive gradient boosting (AGB; AdaBoost). Among these techniques, FCNN yielded remarkable accuracy with 95% accuracy. Ahmed et al.[31] predicted the viscosity of CO_2 foam fracturing, considering different variables: foam quality, shear rate, temperature, and pressure using the deep neural networks approach. The fracturing CO_2 foam viscosity was predicted with 99.8% accuracy and a low average absolute relative deviation (<1.8%). Tariq et al.[32] analyzed 360 data points of effective foam viscosity using four AI techniques: K-nearest neighbor (KNN), artificial neural network (ANN), random forest regressor (RFR), and DTs. The study tested a wide range of conditions using a high-pressure high-temperature foam rheometer and utilizing seven commercial surfactants. The ANN-based model predicted foam viscosity with an optimum accuracy ($R^2 = 0.972$).

Few papers studied the prediction of foam viscosity using machine learning tools due to the complex nature of the foam structure. Additionally, the generation of foam viscosity experimentally at high pressure and high temperature was not trivial due to the limitation of the experimental setup. Compared with previous research, this paper presents an

effective model to predict foam viscosity based on popular rheological models such as the power-law, Bingham plastic, or Casson models. The experimental data were generated using a high-pressure, high-temperature foam rheometer at a range of temperature (100−150 °C), shear rate (100−2000 1/s), pressure (1000−2500 psi), and foam quality of 70%. The effect of different variables was investigated such as surfactant types, concentrations, water chemistry, gas types, chelating agents, and CIs. Machine learning approaches were used to derive a model for foam viscosity based on the rheological model constants, considering the features above. In addition, the significance of each parameter on foam viscosity was investigated through Pearson's correlation coefficients.

2. METHODOLOGY

2.1. Chemicals and Sample Preparation. This paper investigated the rheology of several surfactants: Duomeen TTM, Ethoduomeen T/13, Armogel O, and Ethomeen C/12. They are nonionic surfactants and converted to cationic surfactants once they protonated. They are composed of alkylamine and diamine derivatives manufactured by Nouryon. Additionally, erucamidopropyl hydroxopropylsultaine is a viscoelastic surfactant and commercially named Armovis VES. It is only soluble in saline water. Moreover, CAS 50 is an amphoteric surfactant synthesized from cocamidopropyl hydroxysultaine. The chemistry of the surfactants that were used is shown in Table 1. In addition, a low pH glutamic acid diacetate tetrasodium salt (GLDA) (pH 3.87) was used in this study. The CI is composed of alkyl polyquaternary amine-based and commercially named CP-411T.

Three types of water were synthesized in-house; produced water (PW), seawater (SW), and formation water (FW). The salinity and ion content of each water type are shown in Table 2.

Table 2. Salt Composition of Synthetic Water

chemical formula	produced water, PW (g/L)	seawater, SW (g/L)	formation water, FW (g/L)
NaCl	14.976	41.172	150.446
CaCl$_2$.2H$_2$O	5.649	2.387	69.841
MgCl$_2$.6H$_2$O	2.258	17.644	20.396
NaHCO$_3$	0.476	0.165	0.487
Na$_2$SO$_4$	1.252	6.339	0.518
TDS (ppm)	24611	67708	241688

Each surfactant was typically dissolved in one of the water types (deionized water, PW, SW, or FW) containing GLDA (0 or 15 wt %) and the CI (0 or 1.5 wt %). HCl droplets were added (37.7%) to the batches that did not include GLDA to reduce the pH and ensure surfactant dissolution. The liquid solution was then stirred for 24 h at 450−500 rpm (RPM) using a magnetic stirrer at room conditions?

2.2. HPHT Foam Rheometer. The apparent viscosity of the foam was determined by using an HPHT foam rheometer, as shown in Figure 1. The foam quality was set to be 70% for all experiments. The viscosity measurements were evaluated at a wide range of shear rates (100−2000 1/s). The experiment started with the loading of a liquid solution into the accumulator. Then, 130 mL of the solution was pumped into a flow loop using a Quizix pump. Then, the oven was set to 70 °C. Once the temperature was reached, CO_2 or N_2 was introduced to the recirculation loop to prevent the evaporation

of the liquid at elevated temperatures. Next, the oven was set to the desired experiment temperature. When the experiment temperature was reached, more gas was pumped through the gas booster until the desired test pressure was reached. Then, shearing was applied for enough time to generate foam using a recirculation pump. The generated foam can be visualized through a viewing camera. In addition, the foam quality was controlled manually by adding more liquid or gas. Once the homogeneous foam was produced and the desired foam quality was set, an automated schedule was set to log the measured value while increasing the shear rate from 100 to 2000 1/s using 15−40 min time intervals. The apparent foam viscosity, μ (mPa·s) is defined as the ratio of shear stress, τ (dyne-s/cm^2), to shear rate, $\dot{\gamma}$ (1/s), evaluated using the following equation:

$$\mu = \frac{\tau}{\dot{\gamma}} = \frac{\left(D\frac{\Delta P}{4L}\right)}{\frac{8U}{D}} \tag{1}$$

where L is the length of the loop (cm), ΔP is the pressure difference, D is the loop diameter (cm), and U is the velocity in the loop (cm/s).

2.3. Foam Rheology Models. The flow behavior of foam fluids is classified as non-Newtonian due to the viscosity variation with shear rates. The power-law, Bingham, or Casson models characterize the connection between shear stress or viscosity and the shear rate.

The power-law model was developed with two factors: the consistency index, K (m·pa·s^n), and the flow behavior index, n (dimensionless). These factors were used to develop the following relationship between shear rate ($\dot{\gamma}$) and shear stress (τ) as follows:

$$\tau = K\dot{\gamma}^n \tag{2}$$

When the flow behavior index (n) is equal to one, the flow fluid acts as a Newtonian. Where the viscosity is unaffected by the shear rate. When n is more than one, the fluid is classified as shear thinning (pseudoplastic). When n is less than one, the moving fluid exhibits shear thinning (dilatant fluid).

The Bingham plastic model was established using two parameters: the fluid's yield stress, τ_y (m·Pa), and its plastic viscosity, μ_p (cp).[33] The fluid initially resists flow until shear stress surpasses a certain threshold. After that, the fluid begins to flow, displaying linear Newtonian fluid behavior between the shear stress and shear rate. The shear stress-shear rate connection may be represented as follows:

$$\tau = \tau_y + \mu_p\dot{\gamma} \tag{3}$$

The Casson fluid model, the shear-thinning model, is derived based on two parameters; Casson yield, K_{oc} (m·Pa), and Casson plastic viscosity, K_c (m·Pa·s). The shear stress-shear rate model is explained as the following:

$$\sqrt{\tau} = \sqrt{K_{oc}} + \sqrt{K_c}\sqrt{\dot{\gamma}} \tag{4}$$

The viscosity measurements were carried out using HPHT foam rheometer at foam quality of 70% and shear rates ranging from 100 to 2000 1/s. Several surfactants, different water chemistry (DI, PW, SW, and FW), temperatures (100−150 °C), pressures (1000−4000 psi), and surfactant concentrations (0.5−2 wt %) were evaluated. Additionally, the impact of chelating agent GLDA and CI on the acidic N_2 foam and CO_2 foam were investigated. The supporting data presented the
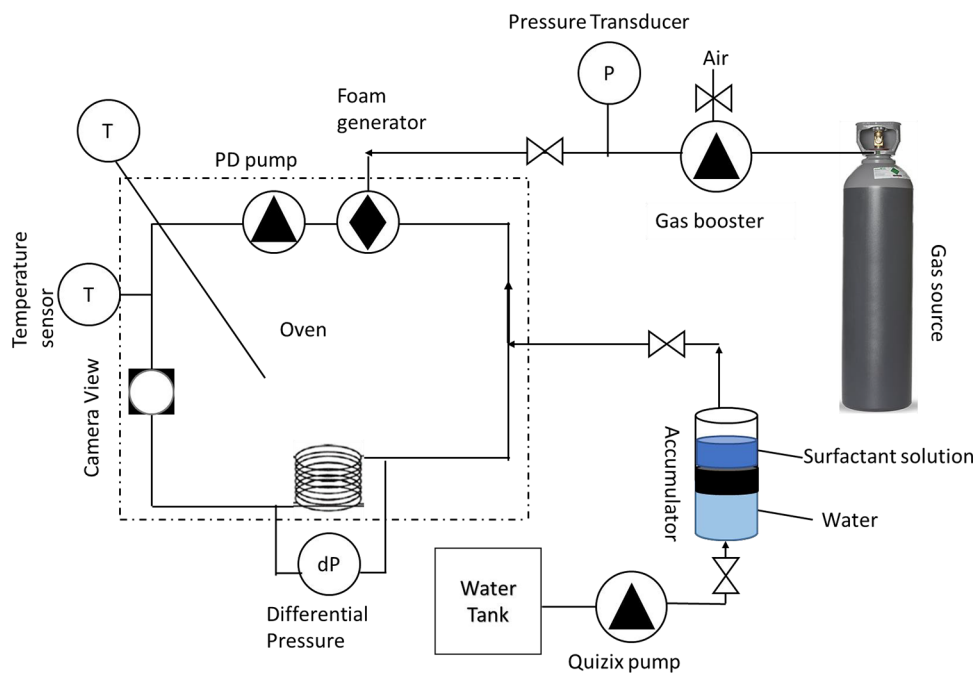
Figure 1. Scheme of the HPHT foam rheometer.

relation between apparent foam viscosity and shear rate. Then, a curve fitting was performed for all experiments to obtain the rheological models' parameters: K_c, K_{oc}, μ_p, τ_y, k, and n using eqs 2−4. The power-law model was the most appropriate model to describe the behavior of the foam with a linear regression coefficient of $R^2 = 0.96$, ranging from 0.99 to 0.79. However, a deviation from the data was observed for some features such as high temperature (150 °C), CI, and Armovis surfactant. The Bingham plastic model was also used to describe the behavior of the foam system. The regression coefficient for all experiments was 0.87, ranging from 0.99 to 0.3561. Deviation from experimental data was detected for many features: salinity, surfactant concentration, surfactant type, gas pressure, and chelating agents. To the best of our knowledge, no publication has investigated the foam flow behavior using the Casson model. This model showed an excellent correlation ($R^2 \sim 0.94$) with the rheological data. Deviation was observed for the gas pressure and gas type.

The generated foam viscosity data were used to extract the rheological models' constants; K_c, K_{oc}, μ_p, τ_y, k, and n using eqs 2−4. The objective of machine learning tools was to predict the rheological models' constants based on features, which were used to evaluate foam viscosity based on the optimum rheological model.

2.4. Data Analysis. In this study, histograms were used to gain insights about the data distributions such as their range, frequency, and the spread. Figure 2 provides a visual representation of the histogram distribution for each parameter used in machine learning predictive modeling.

To examine the degree of collinearity among the input parameters, we constructed a heatmap utilizing Pearson's method. The collinearity refers to a strong linear association between two or more input parameters. Figure 3 illustrates the heatmap representing the correlations among all the parameters.

2.5. Exploratory Data Analysis. Table 3 provides the nomenclature and meanings of the 19 selected features derived

from the experimental data set, which were employed for predicting the foam rheological parameters.

In this study, a total of 448 data points were collected from laboratory experiments. The statistical analysis of the data set was conducted on the obtained data set. The analysis included identifying the mean, minimum, maximum, ranges, kurtosis, skewness, standard deviation, variance, and interquartile ranges. Table 4 lists all of these statistical analyses.

2.6. Performance Metrics. This study used several error metrics to evaluate the predictive performances of the model, including predictive cross-plots, average absolute percentage error (AAPE), R^2, root-mean-square error (RMSE), and SD.

2.6.1. AAPE. AAPE is a metric used to evaluate a prediction model's degree of accuracy. The average absolute percentage difference between the expected and actual values is quantified. The absolute difference between each predicted value and its matching actual value is taken, divided by the actual value, and averaged throughout the data set to determine the AAPE:

$$\text{AAPE} = \frac{1}{n_{\text{samples}}} \sum |(x_{\text{measured}} - x_{\text{pred}})| \times \frac{100}{x_{\text{exp}}} \tag{5}$$

2.6.2. Root-Mean-Square Error. RMSE calculates the square root of the mean of the squared differences between the predicted values and actual values:

$$\text{RMSE} = \sqrt{\frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} (x_{\text{measured}} - x_{\text{pred}})^2} \tag{6}$$

2.6.3. Coefficient of Determination (R^2). The R^2 ranges from 0 to 1, with higher values indicating a better fit of the model to the data:

$$R^2 = \left(\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \right)^2 \tag{7}$$
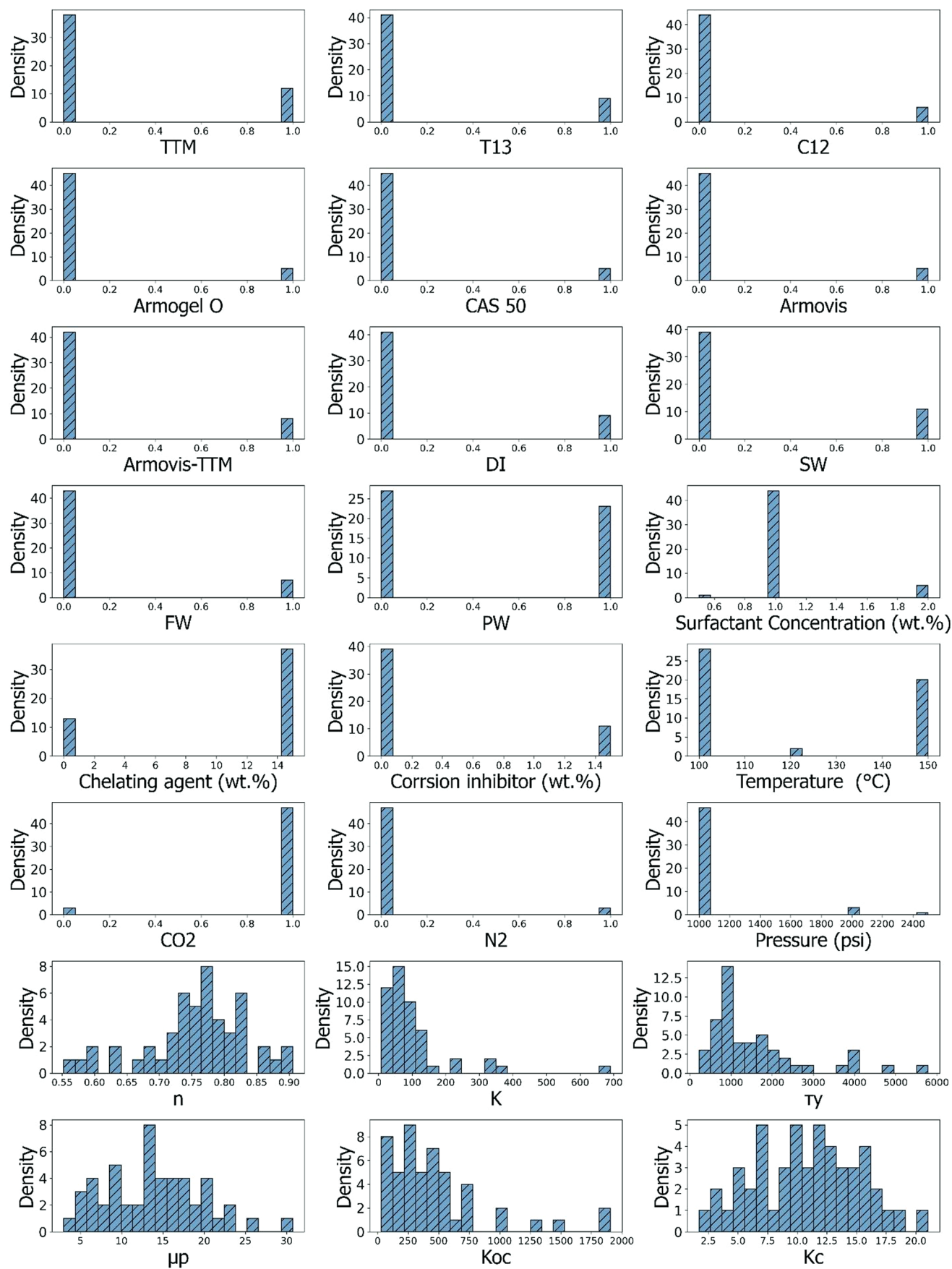
Figure 2. Histograms of input and output parameters used for machine learning models training.
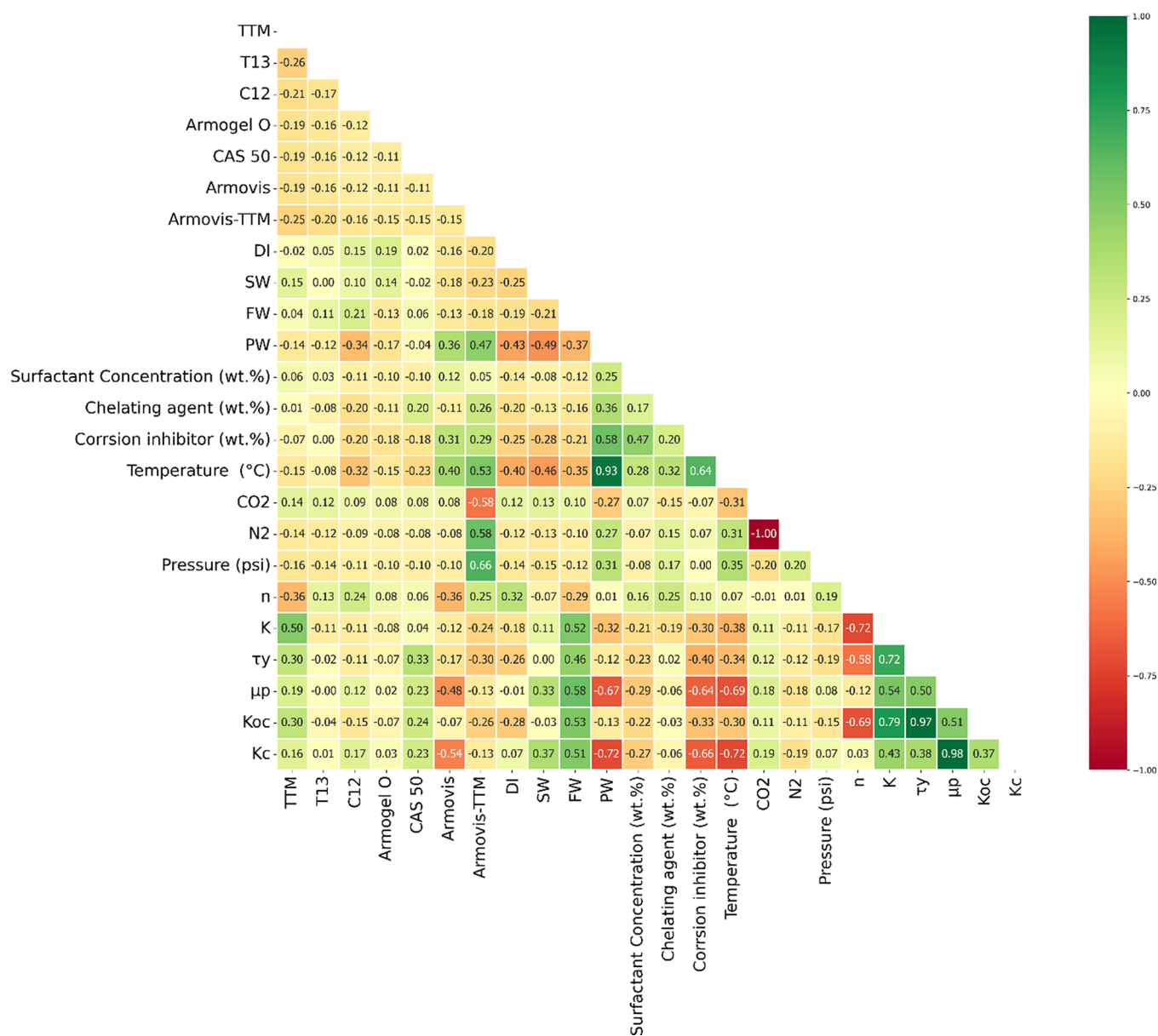
Figure 3. Visualizing the Pearson correlation coefficient between variables across an entire data set through a heatmap.

2.6.4. Standard Deviation. SD is a measure of the dispersion or spread of a data set. It quantifies the extent to which data points deviate from the mean (average) value. A higher standard deviation indicates greater variability in the data, while a lower standard deviation suggests that the data points are closer to the mean:

$$\text{SD} = \sqrt{\frac{1}{n_{\text{samples}} - 1} \sum_{i=1}^{n} \left(\frac{x_{\text{measured}} - x_{\text{pred}}}{x_{\text{measured}}} \right)^2} \qquad (8)$$

2.6.5. Residual Errors. It represents the difference between the observed (actual) values and the predicted values generated by a statistical model. In other words, it quantifies how much the model's predictions deviate from the actual data points:

$$\text{RE} = x_{\text{measured}} - x_{\text{pred}} \qquad (9)$$

where x_{exp} is the actual value and x_{pred} is the predicted value from the model, n_{samples} are the total number of data points, and x and y are the two variables.

2.7. Machine Learning Techniques. In this study, seven different ML techniques were implemented to predict the foam viscosity model based on rheological models such as power-law, Bingham plastic, and Casson models. These seven ML techniques include DT, RF, extreme gradient boosting (XGB), adaptive boosting (AdaBoost), gradient boosting (GBR), support vector regression (SVR), and voting regression (VR).

These ML techniques are mostly used for various tasks, such as classification, regression, and data analysis.[34] Each algorithm has its unique characteristics and advantages, which make it suitable for different types of problems.[35−37]

The DT is a generalized algorithm that creates a flowchart-like structure by partitioning the data based on feature values. Each internal node represents a feature test, and each leaf node represents a class label or regression value. DTs are known for their interpretability and ease of understanding. They can

Table 3. Meaning of Features and Their Nomenclature

feature	type of variable	physical meaning	units
TTM	categorical	presence or absence of TTM surfactant	
T13	categorical	presence or absence of T13 surfactant	
C12	categorical	presence or absence of C12 surfactant	
Armogel O	categorical	presence or absence of Armogel O surfactant	
CAS 50	categorical	presence or absence of CAS 50 surfactant	
Armovis	categorical	presence or absence of Armovis surfactant	
Armovis-TTM	categorical	presence or absence of Armovis-TTM surfactant	
DI	categorical	presence or absence of DI	
SW	categorical	presence or absence of SW	
FW	categorical	presence or absence of FW	
PW	categorical	presence or absence of PW	
CO_2	categorical	presence or absence of CO_2	
N_2	categorical	presence or absence of N_2	
surfactant concentration (wt %)	numerical	weight % of the surfactant used	wt %
chelating agent (wt %)	numerical	weight % of the chelating agent used	wt %
corrosion inhibitor (wt %)	numerical	weight % of the corrosion inhibitor used	wt %
temperature (°C)	numerical	temperature at which measurements were conducted	°C
pressure (psi)	numerical	pressure at which experiments were conducted	psi
n	numerical	Newtonian constant	
K	numerical	consistency index	$m \cdot pa \cdot s^n$
τ_y	numerical	yield stress	$m \cdot Pa$
μ_p	numerical	pseudoplastic viscosity	cp
K_{oc}	numerical	orifice viscosity	$m \cdot Pa$
K_c	numerical	Casson viscosity	$m \cdot Pa \cdot s$

handle both numerical and categorical data and are often used for classification and regression tasks.[38−41]

RF is an ensemble learning algorithm that combines multiple DTs. It creates a collection of trees and makes predictions by averaging the predictions of individual trees. RF improves DTs by reducing overfitting and increasing accuracy. It introduces randomness in the training process by using random subsets of the data and random subsets of the features. RF is robust and performs well on large data sets.[42−45]

SVR is a regression algorithm that builds on the principles of support vector machines. SVR finds a hyperplane that maximizes the margin while still allowing some errors within a specified tolerance (epsilon). It maps the input data into a higher-dimensional space by using a kernel function and constructs the regression model. SVR is robust against outliers and can handle nonlinear regression tasks effectively. It is widely used in various domains, including finance, bioinformatics, and engineering.[46−50]

XGBoost, is a powerful GB algorithm that has gained popularity in machine learning competitions. It is an ensemble method that combines weak learners, typically DTs, in a sequential manner. XGBoost optimizes a differentiable loss function by iteratively adding trees that minimize the loss. It employs various techniques such as regularization, parallel

processing, and weighted quantile sketching for efficient and accurate predictions. XGBoost has become a standard algorithm in many machine learning applications.[51−54]

AdaBoost, short for adaptive boosting, is another popular boosting algorithm that combines weak learners to create a strong learner. It assigns weights to the training instances and adjusts these weights in each iteration to focus on the misclassified samples. AdaBoost places more emphasis on difficult examples, enabling the model to improve its performance iteratively. The final prediction is a weighted combination of weak learners. AdaBoost is robust against overfitting and can be used for both classification and regression tasks.[55−57]

GB is a general boosting framework that combines multiple weak learners to create a strong learner. It minimizes a loss function by iteratively adding new models that fit the negative gradient of the loss function. GB algorithms, such as XGBoost and AdaBoost, sequentially train models to correct the mistakes made by previous models. The final prediction is a weighted sum of the predictions from all of the models. GB algorithms are highly effective and have achieved state-of-the-art results in various machine learning tasks.

VR is an ensemble learning technique, where multiple regression models are combined to make predictions. Each individual regression model in the ensemble provides its own prediction, and the final prediction is determined by aggregating the results from all of the models. This aggregation can be done using different strategies, such as averaging the predictions or taking the majority vote. By combining the predictions from multiple models, VR leverages the diversity of the models to improve prediction accuracy and robustness. It is particularly useful when the individual regression models have different strengths and weaknesses, as the ensemble can compensate for individual model limitations and provide more accurate and reliable predictions.[58−60] Figure 4 shows the architecture of the VR.

3. RESULTS AND DISCUSSIONS

The study's objective is to create machine learning models capable of predicting the foam rheology model parameters. To identify influential factors affecting foam rheology, input parameters were chosen by using laboratory tests. Furthermore, a comprehensive series of numerical experiments was carried out to identify the most effective machine learning models. To convert categorical variables into a numerical form, one-hot encoding was employed. This approach assigns binary values like "0" and "1" to input parameters, where "0" represents the absence of a specific quantity and "1" indicates its presence.

In this research, we leveraged the concept of a learning curve as a valuable tool for evaluating the performance and effectiveness of our machine learning models. The learning curve has allowed us to gain insights into how the model's performance evolves with the increase in the size of the training data set. We visually represented this learning curve by plotting the model's performance in terms of RMSE on the vertical axis against the varying sizes of the training data set on the horizontal axis, as depicted in Figure 5. To create the learning curve, we divided our complete data set into several subsets of different sizes. These subsets were then further divided into a training set, utilized for model training, and a separate validation set, employed to assess the model's performance. We systematically expanded the training set

Table 4. Statistical Analysis of the Entire Dataset

| parameters	average	standard deviation	minimum	25%	50%	75%	maximum
TTM	0.254902	0.440143	0	0	0	0.5	1
T13	0.176471	0.385013	0	0	0	0	1
C12	0.117647	0.325396	0	0	0	0	1
Armogel O	0.098039	0.300327	0	0	0	0	1
CAS 50	0.098039	0.300327	0	0	0	0	1
Armovis	0.098039	0.300327	0	0	0	0	1
Armovis-TTM	0.156863	0.36729	0	0	0	0	1
DI	0.176471	0.385013	0	0	0	0	1
SW	0.215686	0.41539	0	0	0	0	1
FW	0.137255	0.34754	0	0	0	0	1
surfactant concentration (wt %)	0.46	0.503457	0	0	0	1	1
chelating agent (wt %)	1.09	0.314448	0.5	1	1	1	2
corrosion inhibitor (wt %)	11.1	6.646312	0	3.75	15	15	15
temperature (°C)	120.3922	24.32783	100	100	100	150	150
CO_2	0.941176	0.237635	0	1	1	1	1
N_2	0.058824	0.237635	0	0	0	0	1
pressure (psi)	1088.235	311.5427	1000	1000	1000	1000	2500
n	0.75559	0.076854	0.5526	0.7279	0.7663	0.8033	0.9059
K	106.8088	115.9883	8.554606	48.56539	73.62071	113.5579	690.5577
τ_y	1602.799	1197.475	225.09	871.255	1196.1	1921.3	5789.3
μ_p	14.06207	5.902646	2.9943	9.2906	13.983	17.819	30.752
K_{oc}	475.9542	417.8331	31.70816	221.0037	362.4074	545.529	1907.069
K_c	10.9015	4.373568	1.712172	7.301894	10.79188	14.01136	21.0424
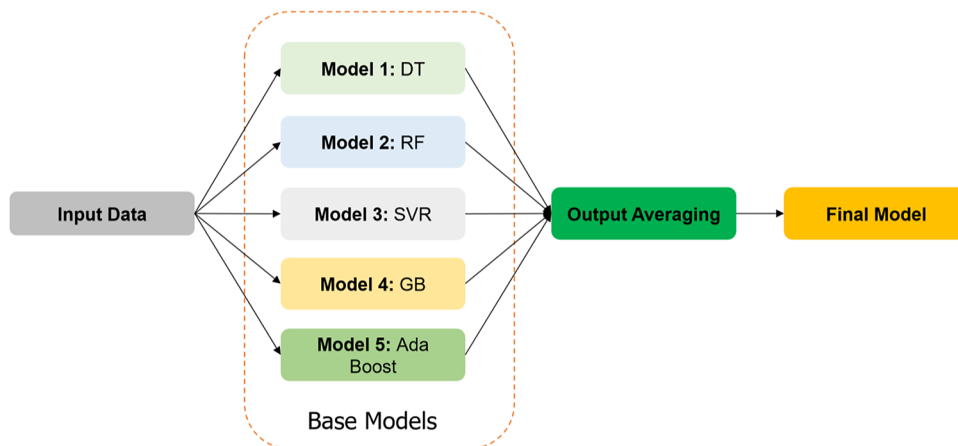
Figure 4. Model architecture of the voting regression.

size and observed how the model's performance responded when applied to the test set. The insights drawn from the learning curve indicated that achieving a training set size of 39 experimental points led to reduced RMSE values in both the training and cross-validation phases, corresponding to 76% of the data set. Consequently, we made t he decision to split the overall data set into three subsets for our machine learning modeling: 76% for training (including 10% for validation) and 24% for testing. To maintain consistency in our modeling approach, we implemented data stratification using the train-test split function from the open-source scikit-learn library in Python. Additionally, we introduced a random seed function to ensure the reproducibility of our results. Furthermore, we applied a cross-validation technique to enhance the accuracy of our machine learning models. This involved dividing the training data set into subsets referred to as "k folds".

We started with the prediction of K_c using these seven ML techniques. The DT was utilized, employing the squared error criterion to determine the quality of splits and allowing the tree to grow without a maximum depth constraint. The RF regressor consisted of 100 DTs with a maximum depth of 100 and a minimum of 2 samples required for splitting. The KNN regressor used 6 neighbors for prediction with uniform weights and employed the Euclidean distance metric. The SVR model employed an RBF kernel with polynomial features up to degree 3, determined the gamma parameter automatically, and set the error tolerance to 0.001. The GBR model included 500 estimators, a learning rate of 0.3, and a maximum depth of 2 for each DT. The AGB model used 500 DTs with a maximum depth of 115, required a minimum of 4 samples for splitting, and had a learning rate of 0.1. XGB employed 500 boosting rounds, minimized the mean squared error, set regularization parameters (λ and α) to control regularization strength, and utilized a gamma value of 0.1. The "max_depth" was set to 100, and the learning rate was 0.3. These models were carefully configured to ensure accurate predictions and capture complex
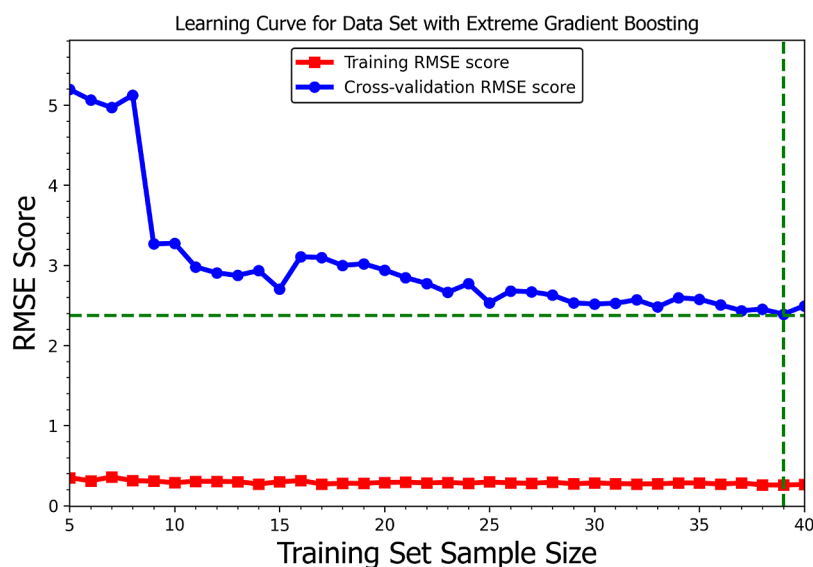
Figure 5. Example of learning curve with increasing training set sample size using XGB.
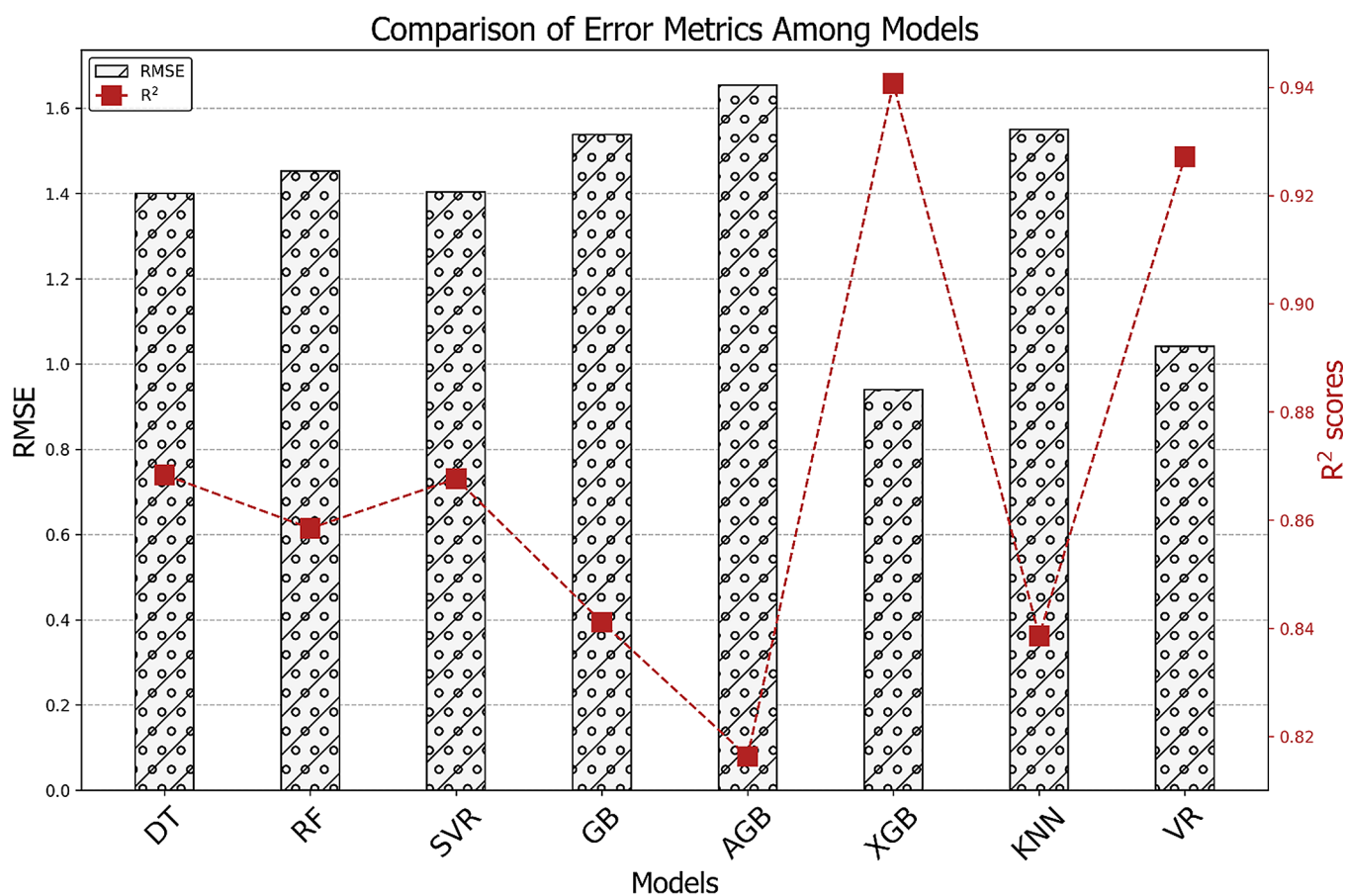

Figure 6. Comparison of different machine learning algorithms in terms of RMSE and R^2 scores to predict K_c.

relationships in the data. The VR employed combines four different models: XGB, AGB, GBR, and RF. Each model contributes to the final prediction by assigning a weight to its output based on its performance. The VR aggregates the predictions from these models and calculates the average as the final prediction. Figure 6 shows the comparison of seven machine learning techniques, including DT, RF, XGB, AGB, GB, SVR, and VR in terms of RMSE and R^2 scores on the

testing data set to predict K_c. The XGB technique outperformed other ML methods by displaying a lower RMSE and higher R^2 values. The XGB resulted in a RMSE of 0.90 and R^2 of 0.941. Figure 7 shows the contribution of each feature in the ML models mentioned above to predict K_c. From Figure 7, it can be observed that almost all features contribute in the XGB model prediction while other ML techniques such as DT, RF, GB, and Adaboost are unequally depending on some features.
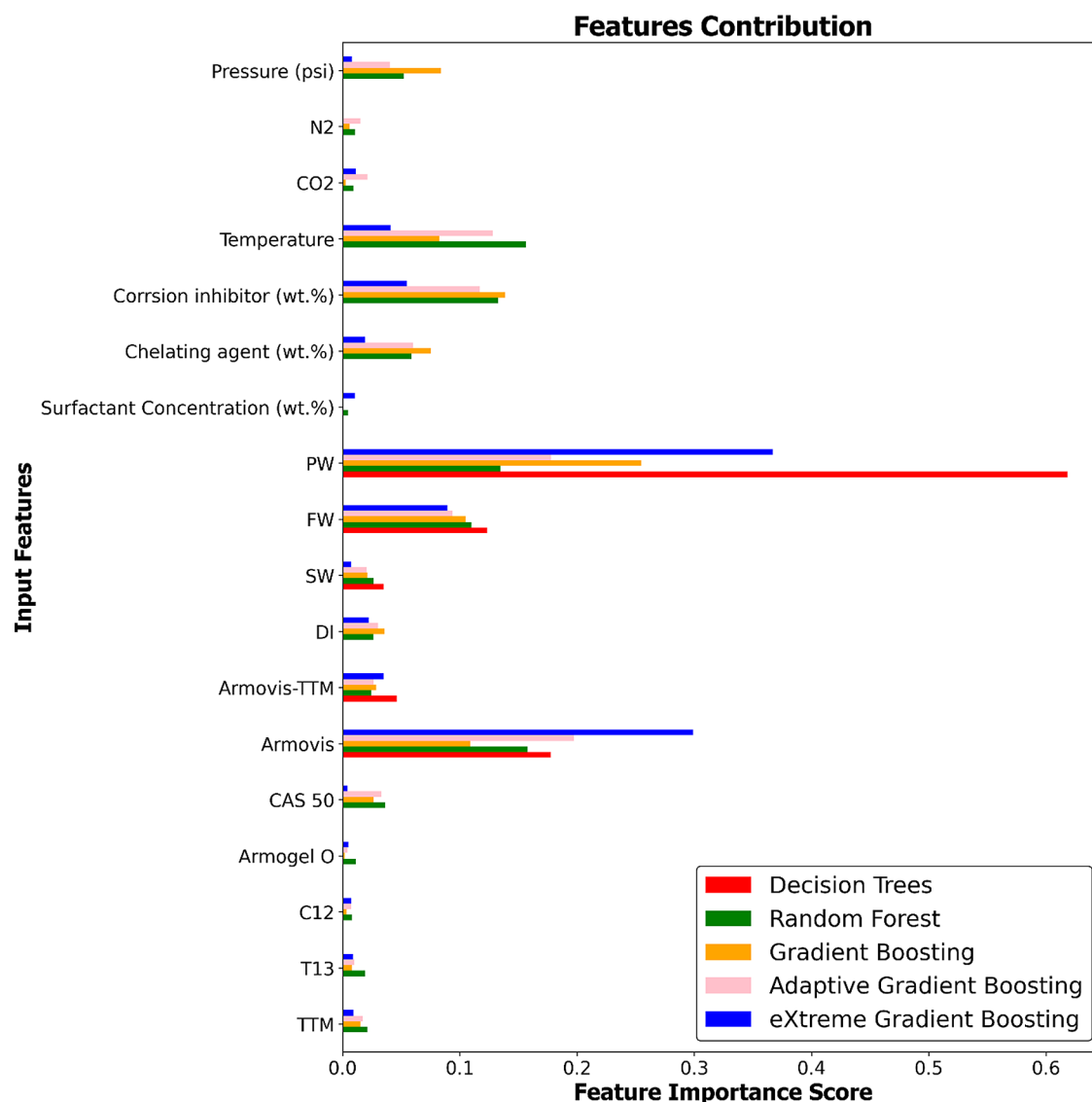
Figure 7. Contributions of each feature in the respective ML models for the prediction of K_c. SVR, KNN, and VR do not support feature importance measuring criteria.

The contributions from each feature are making the XGB model more generalized, therefore, XGB was opted for the prediction of the rest of the rheological parameters.

Using the above configuration of XGB, Figure 8a–f shows the training and testing cross-plots for the prediction of K_c, K_{oc}, μ_p, τ_y, k and n. Table 5 provides information on the training and testing accuracies of various rheological parameters using the XGBoost algorithm. In the "Training" section of the table, the values represent the accuracies of the model when trained on a particular data set. The metrics used to evaluate the model's performance include AAPE, RMSE, MSE, Emax, Emin, and R^2. Each variable has corresponding values for each of these metrics. The XGB model achieves an AAPE of 0.022 for K_c, 0.001 for K_{oc}, 0.020 for μ_p, 0.000 for τ_y, 0.004 for k, and 0.012 for n. Similarly, the model achieves corresponding values for MAE, RMSE, MSE, E_{max}, E_{min}, and R^2 for each variable during the training and testing phases. These values provide insights into the performance of the XGBoost model in predicting the rheological parameters. The lower the values for AAPE, MAE, RMSE, and MSE, the better the accuracy of the model. Conversely, higher values for R^2 indicate a better fit of

the model to the data. It is important to compare the training and testing accuracies to assess the model's ability to generalize to new, unseen data.

4. CONCLUSIONS

Based on the findings and discussions outlined in the paper, the following conclusions can be drawn:

1. The study explored seven distinct machine learning methodologies, including DT, RF, XGB, AdaBoost, GBR, SVR, and VR. The objective was to enhance the accuracy of predicting rheological properties. These models were trained using operational features such as pressure, temperature, and the proportions of various additives.

2. Utilizing machine learning techniques, this study effectively automated the prediction of rheological properties using parameters extracted from labor-intensive and time-consuming laboratory experiments. These techniques enabled the development of effective models for predicting rheological features.
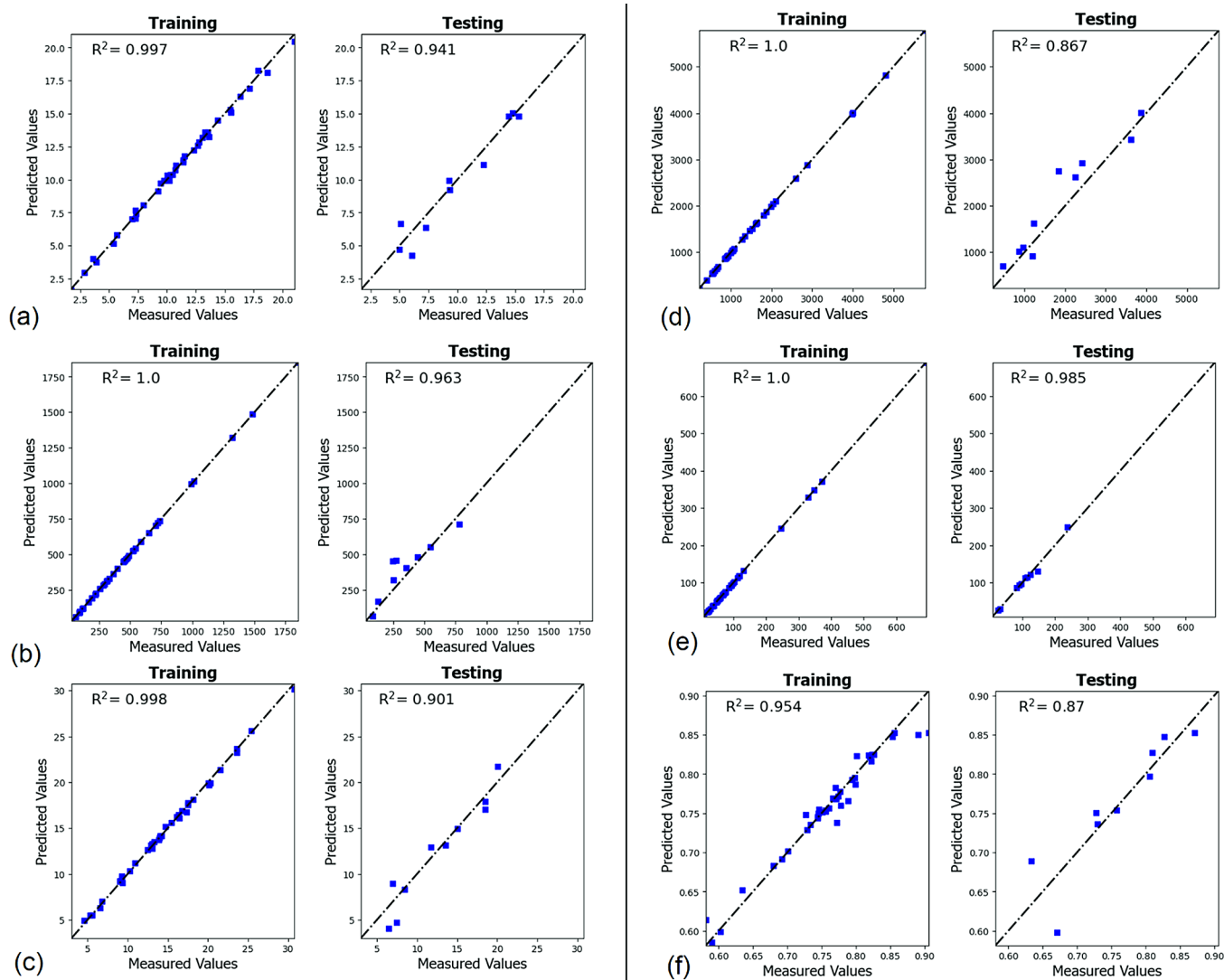
Figure 8. Training and testing predictions using XGB for (a) K_c, (b) K_{oc}, (c) μ_p, (d) τ_p, (e) k, and (f) n.

Table 5. Training and Testing Accuracies of the Rheological Parameters Using XGB

variables	AAPE	MAE	RMSE	MSE	E_{max}	E_{min}	R^2
(a) training							
K_c	0.022	0.202	0.256	0.066	0.572	−0.434	0.997
K_{oc}	0.001	0.178	0.229	0.052	0.687	−0.373	1.000
μ_p	0.020	0.222	0.268	0.072	0.621	−0.484	0.998
τ_y	0.000	0.257	0.304	0.093	0.621	−0.587	1.000
k	0.004	0.210	0.280	0.078	0.744	−0.728	1.000
n	0.012	0.009	0.015	0.000	0.053	−0.032	0.954
(b) testing							
K_c	0.104	0.761	0.940	0.883	1.793	−1.590	0.941
K_{oc}	0.266	73.348	97.418	9490.250	67.134	−206.304	0.963
μ_p	0.139	1.284	1.557	2.423	2.700	−2.028	0.901
τ_y	0.234	332.260	403.171	162547.048	281.459	−913.532	0.867
k	0.039	4.445	6.999	48.986	17.236	−12.320	0.985
n	0.038	0.026	0.033	0.001	0.073	−0.055	0.870

3. A substantial volume of data was meticulously generated through extensive laboratory experiments, forming the cornerstone for training the machine learning models. This data set provided the necessary groundwork for the models to discern the inherent patterns and connections

between input parameters and the foam rheological parameters.

4. The outcomes of the study revealed that the XGB model surpassed the other tested algorithms in terms of predictive performance. This highlights the XGB

model's exceptional capability in capturing the intricate relationships between input parameters and output parameters.

■ AUTHOR INFORMATION

Corresponding Authors

Murtada Saleh Aljawad — *College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum and Minerals (KFUPM), Dharahan 31261, Saudi Arabia; Center for Integrative Petroleum Research, King Fahd University of Petroleum and Minerals, 31261 Dhahran, Saudi Arabia;* orcid.org/0000-0002-3540-6807; Email: mjawad@kfupm.edu.sa

Zeeshan Tariq — *Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia;* Email: zeeshan.tariq@kaust.edu.sa

Authors

Jawad Al-Darweesh — *College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum and Minerals (KFUPM), Dharahan 31261, Saudi Arabia*

Shabeeb Alajmei — *College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum and Minerals (KFUPM), Dharahan 31261, Saudi Arabia; Center for Integrative Petroleum Research, King Fahd University of Petroleum and Minerals, 31261 Dhahran, Saudi Arabia;* orcid.org/0000-0002-5816-4214

Bicheng Yan — *Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955, Saudi Arabia;* orcid.org/0000-0002-3356-7594

Muhammad Shahzad Kamal — *College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum and Minerals (KFUPM), Dharahan 31261, Saudi Arabia; Center for Integrative Petroleum Research, King Fahd University of Petroleum and Minerals, 31261 Dhahran, Saudi Arabia;* orcid.org/0000-0003-2359-836X

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsomega.4c00965

Author Contributions

The manuscript was written through contributions of all authors.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Aljawad, M. S.; Aljulaih, H.; Mahmoud, M.; Desouky, M. Integration of field, laboratory, and modeling aspects of acid fracturing: A comprehensive review. *J. Pet. Sci. Eng.* **2019**, *181*, 106158.

(2) Rahim, Z.; Al-Kanaan, A. A.; Waheed, A.; Kayumov, R.; Al-Jalal, Z. Foamed Acid Fracturing Enables Efficient Stimulation of Low-Pressure Carbonate Reservoirs. In *SPE International Hydraulic Fracturing Technology Conference and Exhibition*; SPE, 2018, p. D023S006R001.

(3) Rahim, Z. Technology Focus: Hydraulic Fracturing (March 2018). *J. Pet. Technol.* **2018**, *70* (03), 58.

(4) Aljawad, M. S.; Schwalbert, M. P.; Zhu, D.; Hill, A. D. Guidelines for Optimizing Acid Fracture Design Using an Integrated Acid Fracture and Productivity Model. In *SPE International Hydraulic Fracturing Technology Conference and Exhibition*; SPE, **2018**, p. D023S006R002.

(5) Wanniarachchi, W. A. M.; Ranjith, P. G.; Perera, M. S. A.; Lashin, A.; Al Arifi, N.; Li, J. C. Current opinions on foam-based hydro-fracturing in deep geological reservoirs. *Geomech. Geophys. Geo-Energy Geo-Resources* **2015**, *1* (3–4), 121–134.

(6) Al-Darweesh, J.; Aljawad, M. S.; Al-Ramadan, M.; Elkatatny, S.; Mahmoud, M.; Patil, S. Review of underbalanced drilling techniques highlighting the advancement of foamed drilling fluids. *J. Pet. Explor. Prod. Technol.* **2022**, *13*, 929.

(7) Karadkar, P.; Harbi, B.; Malik, A.; Alsakkaf, M.; Khan, S. CO2 Foamed Fracturing Fluids for High Temperature Hydraulic Fracturing. In *Middle East Oil, Gas and Geosciences Show*; SPE, 2023, p. D011S019R003.

(8) Obisesan, O.; Ahmed, R.; Amani, M. The effects of oil contamination on the stability and flow behavior of drilling foams under high-pressure conditions. *Results Eng.* **2023**, *17*, No. 100885.

(9) Sinha, V.; Ahmed, R.; Akhtar, T.; Shah, S.; Amani, M. Rheology and hydraulics of polymer-based foams at elevated temperatures. *J. Pet. Sci. Eng.* **2019**, *180*, 330–346.

(10) Sheng, Y.; Zhang, H.; Ma, L.; Wang, Z.; Hu, D.; Zhang, S. Rheological Properties of Gel Foam Co-Stabilized with Nanoparticles, Xanthan Gum, and Multiple Surfactants. *Gels* **2023**, *9* (7), 534.

(11) Fu, C.; Liu, N. Rheology and stability of nanoparticle-stabilized CO2 foam under reservoir conditions. *J. Pet. Sci. Eng.* **2021**, *196*, No. 107671.

(12) Ahmed, R. M.; et al. Rheology of foamed cement. *Cem. Concr. Res.* **2009**, *39* (4), 353–361.

(13) Firoze Akhtar, T.; Ahmed, R.; Elgaddafi, R.; Shah, S.; Amani, M. Rheological behavior of aqueous foams at high pressure. *J. Pet. Sci. Eng.* **2018**, *162*, 214–224.

(14) Sherif, T.; Ahmed, R.; Shah, S.; Amani, M. Rheological correlations for oil-based drilling foams. *J. Nat. Gas Sci. Eng.* **2016**, *35*, 1249–1260.

(15) Einstein, A. Effect of suspended rigid spheres on viscosity. *Ann. Phys.* **1906**, *19*, 289–306.

(16) Mooney, M. The viscosity of a concentrated suspension of spherical particles. *J. Colloid Sci.* **1951**, *6* (2), 162–170.

(17) Frankel, N. A.; Acrivos, A. On the viscosity of a concentrated suspension of solid spheres. *Chem. Eng. Sci.* **1967**, *22* (6), 847–853.

(18) Princen, H. M.; Kiss, A. D. Rheology of foams and highly concentrated emulsions: IV. An experimental study of the shear viscosity and yield stress of concentrated emulsions. *J. Colloid Interface Sci.* **1989**, *128* (1), 176–187.

(19) Al-Darweesh, J.; et al. Investigation of amine-based surfactants for foamed acid stimulation at high temperature, pressure, and salinity. *Geoenergy Sci. Eng.* **2023**, *229*, No. 212094.

(20) Al-Darweesh, J.; Aljawad, M. S.; Al-Yousif, Z.; BinGhanim, A.; Kamal, M. S.; Mahmoud, M. The Impact of Green Chelating Agent and Corrosion Inhibitor on Foam Rheology and Stability at High Temperature, Pressure, and Salinity. *SPE J.* **2022**, *28*, 1216–1229.

(21) Al-Darweesh, J.; Aljawad, M. S.; Kamal, M. S.; Al-Yousef, Z.; Mahmoud, M. Enhancing CO2 Foam Viscosity and Stability at Harsh Reservoir Conditions Through the Synergetic Interactions of Surfactant and Polymer. In *Abu Dhabi International Petroleum Exhibition and Conference*; SPE, 2022, p. D041S129R001.

(22) Kadafur, I.; BinGhanim, A.; Aljawad, M. S.; Kamal, M. S.; AlYousef, Z.; Mahmoud, M. Rheological study of CO2 foamed chelating stimulation fluids under harsh reservoir conditions. *J. Pet. Sci. Eng.* **2022**, *208*, No. 109201.

(23) Ibrahim, A. F.; Nasr-El-Din, H. A. *CO2 Foam for Enhanced Oil Recovery Applications*, Xu, H.; Yang, C.; Jing, D., Eds.; IntechOpen: Rijeka, 2019; Ch. 7.

(24) Al-Darweesh, J.; Aljawad, M. S.; Kamal, M. S.; Mahmoud, M.; Al-Yousef, Z.; Al-Shehri, D. Water chemistry role in the stability of

CO_2 foam for carbon sequestration in water aquifers. *Gas Sci. Eng.* **2023**, *118*, No. 205090.

(25) Wang, Y.; et al. The stability study of CO_2 foams at high pressure and high temperature. *J. Pet. Sci. Eng.* **2017**, *154*, 234−243.

(26) Almobarky, M.; AlYousef, Z.; Schechter, D. Enhancing the Foam Stability Using Surfactants Mixtures. In *SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition*; SPE, 2018, SPE-192449-MS.

(27) Nasr-El-Din, H. A.; Al-Ghamdi, A. H.; Al-Qahtani, A. A.; Samuel, M. M. Impact of Acid Additives on the Rheological Properties of a Viscoelastic Surfactant and Their Influence on Field Application. *SPE J.* **2008**, *13* (01), 35−47.

(28) Yu, W.; Kanj, M. Y. Review of foam stability in porous media: The effect of coarsening. *J. Pet. Sci. Eng.* **2022**, *208*, No. 109698.

(29) AlYousef, Z.; Schechter, D. The Synergy of Surfactant and Nanoparticles: Towards Enhancing Foam Stability. In *SPE Kuwait Oil & Gas Show and Conference*; SPE, 2019, p. D023S007R002.

(30) Othman, A.; Tariq, Z.; Aljawad, M. S.; Yan, B.; Kamal, M. S. Enhancing Fracturing Fluid Viscosity in High Salinity Water: A Data-Driven Approach for Prediction and Optimization. *Energy Fuels* **2023**, *37* (17), 13065−13079.

(31) Ahmed, S.; Alameri, W.; Ahmed, W. W.; Khan, S. A. Rheological behavior of scCO2-Foam for improved hydrocarbon recovery: Experimental and deep learning approach. *J. Pet. Sci. Eng.* **2021**, *203*, No. 108646.

(32) Tariq, Z.; BinGhanim, A.; Aljawad, M. S.; Kamal, M. S.; Mahmoud, M.; AlYousef, Z. AI-driven foam rheological model based on HPHT foam rheometer experiments. *J. Pet. Sci. Eng.* **2022**, *213*, No. 110439.

(33) Faroughi, S. A.; Pruvot, A. J.-C. J.; McAndrew, J. The rheological behavior of energized fluids and foams with application to hydraulic fracturing: Review. *J. Pet. Sci. Eng.* **2018**, *163*, 243−263.

(34) Tariq, Z.; Murtaza, M.; Mahmoud, M.; Aljawad, M. S.; Kamal, M. S. Machine learning approach to predict the dynamic linear swelling of shales treated with different waterbased drilling fluids. *Fuel* **2022**, *315*, No. 123282.

(35) Desouky, M.; Tariq, Z.; Aljawad, M. S.; Alhoori, H.; Mahmoud, M.; Abdulraheem, A. Machine Learning-Based Propped Fracture Conductivity Correlations of Several Shale Formations. *ACS Omega* **2021**, *6* (29), 18782−18792.

(36) Desouky, M.; Tariq, Z.; Aljawad, M. S.; Alhoori, H.; Mahmoud, M.; AlShehri, D. Data-Driven Acid Fracture Conductivity Correlations Honoring Different Mineralogy and Etching Patterns. *ACS Omega* **2022**, *5* (27), 16919−16931.

(37) Tariq, Z.; et al. Machine Learning-Based Accelerated Approaches to Infer Breakdown Pressure of Several Unconventional Rock Types. *ACS Omega* **2022**, *7* (45), 41314−41330.

(38) Kingsford, C.; Salzberg, S. L. What are decision trees? *Nat. Biotechnol.* **2008**, *26* (9), 1011−1013.

(39) Sun, R.; Wang, G.; Zhang, W.; Hsu, L. T.; Ochieng, W. Y. A gradient boosting decision tree based GPS signal reception classification algorithm. *Appl. Soft Comput.* **2020**, *86*, No. 105942.

(40) Kotsiantis, S. B. Decision trees: a recent overview. *Artif. Intell. Rev.* **2013**, *39* (4), 261−283.

(41) Somvanshi, M.; Chavan, P.; Tambade, S.; Shinde, S. V. A review of machine learning techniques using decision tree and support vector machine. In *2016 International Conference on Computing Communication Control and automation (ICCUBEA)*; IEEE, 2016; pp. 1−7.

(42) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45* (1), 5−32.

(43) Svetnik, V.; Liaw, A.; Tong, C.; Christopher Culberson, J.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43* (6), 1947−1958.

(44) Granitto, P. M.; Furlanello, C.; Biasioli, F.; Gasperi, F. Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemom. Intell. Lab. Syst.* **2006**, *83* (2), 83−90.

(45) Aulia, A.; Rahman, A.; Quijano Velasco, J. J. Strategic Well Test Planning Using Random Forest. In *SPE Intelligent Energy International Conference and Exhibition*; SPE, 2014.

(46) Drucker, H.; Surges, C. J. C.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. In *Advances in Neural Information Processing Systems*; MIT Press, 1997; Vol. 9, pp. 155-161.

(47) Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20* (3), 273−297.

(48) Trontl, K.; Šmuc, T.; Pevec, D. Support vector regression model for the estimation of γ-ray buildup factors for multi-layer shields. *Ann. Nucl. Energy* **2007**, *34* (12), 939−952.

(49) Khan, M. R.; Alnuaim, S.; Tariq, Z.; Abdulraheem, A. Machine learning application for oil rate prediction in artificial gas lift wells. In *SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings*; SPE, 2019.

(50) El-Sebakhy, E.; Sheltami, T.; Al-Bokhitan, S.; Shaaban, Y.; Raharja, I.; Khaeruzzaman, Y. Support vector machines framework for predicting the PVT properties of crude-oil systems. In *SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings*; SPE, 2007; Vol. 3, pp. 1416-1429.

(51) Tariq, Z.; Gudala, M.; Yan, B.; Sun, S.; Mahmoud, M. A fast method to infer Nuclear Magnetic Resonance based effective porosity in carbonate rocks using machine learning techniques. *Geoenergy Sci. Eng.* **2023**, *222*, No. 211333.

(52) Wang, C.; Deng, C.; Wang, S. Imbalance-XGBoost: leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. *Pattern Recognit. Lett.* **2020**, *136*, 190−197.

(53) Wu, Y.; Tao, J.; Wang, J.; Zhang, Y.; Peng, S. Experimental investigation of shale breakdown pressure under liquid nitrogen pre-conditioning before nitrogen fracturing. *Int. J. Min. Sci. Technol.* **2021**, *31*, 611.

(54) Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, 13−17-Aug, pp. 785−794.

(55) Amar, M. N.; Shateri, M.; Hemmati-Sarapardeh, A.; Alamatsaz, A. Modeling oil-brine interfacial tension at high pressure and high salinity conditions. *J. Pet. Sci. Eng.* **2019**, *183*, No. 106413.

(56) Schapire, R. E. Explaining adaboost. *Empir. Inference Festschrift Honor Vladimir N. Vapnik* **2013**, 37−52.

(57) Wang, S.; Chen, S. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. *J. Pet. Sci. Eng.* **2019**, *174*, 682.

(58) Geng, Z.; et al. Predicting seismic-based risk of lost circulation using machine learning. *J. Pet. Sci. Eng.* **2018**, *176*, 679−688.

(59) Oñate, E. et al. *Advances in computational plasticity a book in honour of D. Roger J. Owen*; Springer, 2018.

(60) Zha, W.; Li, X.; Xing, Y.; He, L.; Li, D. Reconstruction of shale image based on Wasserstein Generative Adversarial Networks with gradient penalty. *Adv. Geo-Energy Res.* **2020**, *4* (1), 107−114.