Prediction of Interfacial Tension Using Machine Learning: A Review of Applied Techniques in Petrochemical/Reservoir Engineering

OKON, JOHNE[1], TINUOLA UDOH[2], BLESSED EMENKA[3]
[1, 2, 3] *Department of chemical/Petrochemical Engineering, Akwa Ibom State University (AKSU), Ikot Akpaden Mkpat Enin, Akwa Ibom State*

Abstract- ***As efforts to increase oil reserves through enhanced oil recovery projects increase globally, interfacial tension in crude oil - brine systems is becoming increasingly significant. In porous media, displacement processes and multi-phase flow are directly impacted by interfacial tension. It has an impact on oil field emulsions' behavior as well. The majority of documented two-phase flow and displacement procedures executed under varying interfacial tensions have been executed for either water-gas, oil-water or oil-gas two-phase systems. One significant factor influencing the displacements of water/oil and water/oil/gas is the interfacial tension between crude oil and brine. Adhesion tension, capillary pressure, capillary number, and the dimensionless time for imbibition are all influenced by interfacial tension. A crucial physical characteristic that influences several processes in the oil and gas sector, including enhanced oil recovery, multi-phase flow, and emulsion stability, is interfacial tension (IFT). Increasing the efficiency and optimizing these processes depend on accurate IFT prediction. This article reviews the various techniques that have been applied to make interfacial tension (IFT) predictions in liquid-liquid and liquid-vapour systems, exploring advanced machine learning (ML) techniques in terms of the variables used for modelling, variable relevance, internal parameters tuning, performance analysis and future prospects of the most advanced algorithms. The Gradient Boosting (GB), Elastic Net Regression (EN), AdaBoost, SVR, CatBoost, and XGBoost algorithms are explored and results of their application from different studies compared.***

Indexed Terms- ***Interfacial Tension, Machine Learning, Algorithm, Boosting***

I. INTRODUCTION

In petrochemical and reservoir engineering, different machine learning technologies, such as particle swarm optimization (PSO), genetic programming (GP), artificial neural networks (ANN), imperialist competitive algorithms (ICA), and generalized regression neural networks (GRN), have been presented in the literature for IFT estimation. (Hosseinzadeh et al., 2016; Li and Misra, 2017; Ameli et al., 2018; Hemmati-Sarapardeh et al., 2018; Amar et al., 2018).Lab techniques are the most effective means of determining the IFT between a hydrocarbon and a surfactant. The weight of drop method, pendant drop, spinning drop, etc. are laboratory techniques used to measure IFT. Others include the molecular-level theories, like the Gibbs adsorption equation and the Young-Laplace equation, which connect IFT to interfacial curvature and the distribution of molecules at the interface (Tadros et al, 2013;Periera et al, 2017; kim et al, 2022;), and computational techniques, like density functional theory (DFT) and Monte Carlo methods, which shed light on the behavior of molecules at the interface (Zhao, 2018; Singh, 2023) are included in theoretical approaches. Experimental methods include the capillary rise method, the spinning drop method, the Wilhelmy plate and Du Noüy ring methods, the pendant and sessile drop techniques, and others that measure the force needed to deform or separate a fluid interface. Other methods involve examining the form and behavior of fuid droplets, such as drop shape analysis and interfacial rheology.(Clegg, C, 2013). One of the drawbacks and difficulties of laboratory techniques is their time-consuming nature. This method is expensive when taking into account the cost of the tests and the chemicals required to perform the IFT test. Thus, it can be very appealing

and useful to develop a model for predicting the IFT between surfactants and hydrocarbons. Studies in the past have discussed how surfactants affect the surface equation of state at the interfacial boundary between two fluids.

A measurement of the force per unit length acting on the boundary between two immiscible fluids is called interfacial tension, or IFT. Normally, dynes/cm (Bui, T. et al. 2021) are used to measure it. IFT is influenced by the fluids' characteristics and compositions as well as the ambient temperature and pressure. For the purpose of precisely forecasting interfacial behavior and streamlining procedures in a variety of applications, it is essential to comprehend how these factors affect IFT. IFT finds widespread application in the oil and gas sector, where it has a major impact on various aspects including enhanced oil recovery, pipeline transportation, gas injection, emulsion stability, acidizing, and carbon capture and storage. In many of these procedures, the IFT between the gas-water and oil-water systems is a crucial factor that impacts the effectiveness, performance, and security of the activities. Lowering the IFT between injected fluids and crude oil, for instance, improves displacement efficiency in enhanced oil recovery and raises oil production (Kalam, S.2023). In pipeline transportation, effective hydrocarbon transportation depends on controlling IFT to avoid problems like emulsions and corrosion. IFT also affects the stability of the emulsion during processing, which influences the effectiveness of separation.(Garmsiri, H. *et al.*2023; Shafei, M. et al 2021; Kalatehno, J. 2021).

II. MODELLING FRAMEWORKS

Developing machine learning models to predict crude oil brine IFT can be achieved via several approaches in terms of the data used in training and testing the model, the type of algorithm implemented as well as the evaluation techniques employed to ensure accuracy. This is so because crude oil/brine IFT is affected by several parameters depending on how they a varied for different reservoirs. As evident in literature, crude oil, brine and surfactant properties can be used to develop ML models to predict IFT in addition to several other parameters depending on the developers aim and objectives of research.

Amar et al, 2019 utilized interfacial tension data of oil/brine systems, pressure (P), total acid number (TAN), temperature (T), the brine solution pH, oil specific gravity (SG), and equivalent salinity of NaCl (Seq) - The equivalent salinity of collected data in different oil samples ranged from 0 to 300000 ppm in various pressures and temperatures. for the modelling. This prediction model is based entirely on the crude oil and brine properties without consideration of the reservoir rock properties which from studies have a critical effect IFT as it interacts with the fluids (Samira et al, 2021). They implemented the GBDT and the AdaBoost algorithms by developing two models for my varying the input parameters. the SVR technique was applied during the modelling to optimize the loss function. The developed models were evaluated using the AARD (Average Absolute Relative Deviation), R^2 (Coefficient of Determination), RMSE (Root Mean Square Error) and SD (Standard Deviation) statistical metrics to determine the best performing model.

The combination of machine learning techniques together with molecular dynamics (MD) simulation was explored by Kirch et al, 2020 to model brine-oil interfacial tension. MD simulations were performed for mono- and multicomponent (toluene, heptane, Heptol, light, and medium) oil systems interfaced with sulfate and chloride brines with varying cations (Na+ , K+ , Ca2+, and Mg2+) and salinity concentration. The molecular model was made up of 36 constituents, a bulk density of 0.84 g/cm^3, and up to 20 carbon atoms per molecule. The organic molecules in the oil model are separated into three distinct fractions based on the PNA approach. The paraffinic (alkyl and branched alkanes) P, the naphthenic (cycloalkanes) N, and the aromatic (A) are the three types.As shown in Figure 1a, the aqueous and oil phases were assembled into 80 × 80 × 160 (Å3) simulation boxes to create the oil/brine interface model. Using the PACKMOL package, all molecules and ions were added at random orientations and positions by adhering to bulk density values (Figure 1b). 37 The systems contained between 95,000 and 110,000 atoms in total. (Kirch et al, 2020)

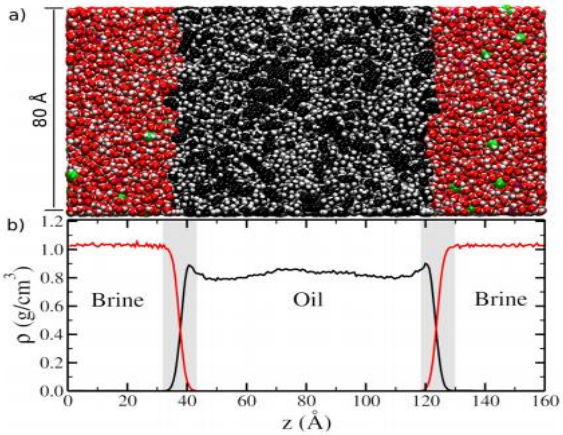

Figure 1: (a) Molecular representation of the oil/brine interface and
(b) the corresponding density profiles for seawater (red) and the
medium-weight oil (black line) models, highlighting the interfaces.
(Kirch et al, 2020)

The Molecular dynamics simulations were performed with the Large Atomic/Molecular Massively Parallel Simulator (LAMMPS) software (plimpton et al, 1995) with accelerator packages for Graphics Processing Unit (GPU). (Brown et al, 2012; Brown et al, 2013). for the machine learning aspect, 130 entries were fed into the ML data set by the obtained IFTs. The feature space is comprised of the oil fraction (P, N, and A), density, brine salinity, and ionic composition (Na+, Ca2+, Mg2+, Cl−, K+, and SO4 2−). Five estimators were taken into consideration when the Scikit-Learn package was applied (Pedregosa et al, 2011): (I) Linear Regression (LR), (II) Random Forest (RF), (III) Extra Trees (ET), (IV) Gradient Boosted (GB), and (V) Elastic Net Regression (EN).

Ali et al. (2023) used reliable tree-based machine learning algorithms to model the interfacial tension of surfactant-hydrocarbon systems. This is an additional intriguing framework that uses data on surfactant (Table 2.0) characteristics to power its modeling. Here, Ali employed the following independent variables for the models: phase inversion temperature (PIT), hydrophilic-lipophilic balance (HLB), surfactant concentration, temperature, and normal alkane molecular weight. This is an enhanced oil recovery (EOR) approach which is quite different from what Amar et al (2019) implemented with regards to IFT prediction. Ali et al, implemented the decision tree (DT), Extra Tree (ET) and the Gradient Boosted Regression Tree (GBRT) algorithms to predict the dependent variable (I.e IFT). similar to Amar (2019), the models accuracy were evaluated with the AARD and R^2 parameters and variable ranking done to ascertain the input variable which has the greatest effect on the model using matplotlib.

Table 2.0: Types and Characteristics of Surfactants used

Surfactant		Chemical formula	dPIT/dx	HLB
Decyl trimethyl ammonium bromide	C_{10}TAB	$C_{10}H_{21}N(CH_3)_3Br$	338	21
Dodecyl trimethyl ammonium bromide	C_{12}TAB	$C_{12}H_{25}N(CH_3)_3Br$	486	19
Myristyl trimethyl ammonium bromide	C_{14}TAB	$C_{14}H_{29}N(CH_3)_3Br$	453	18
Hexadecyl trimethyl ammonium bromide	C_{16}TAB	$C_{16}H_{33}N(CH_3)_3Br$	426	17
Sodium dodecyl sulfate	SDS	$C_{12}H_{25}NaSO_4$	499	40

Seddon et al, (2022) implemented a machine learning hybrid approach to predict surface tension of hydrocarbon surfactants in acqeous solution. The Szyszkowski equation is fitted to a dataset of SFT for 154 model hydrocarbon surfactants at 20–30C to extract three characteristic parameters (Cmax, K_L,

and critical micelle concentration, or CMC), which are correlated to a number of 2D and 3D molecular descriptors (Table 2.1). By eliminating co-correlation and using a gradient-boosted regressor model to rank feature importance and perform recursive feature elimination (RFE), key descriptors were chosen. Using a randomised cross-validated grid search, the hyper-parameters of each target-variable model were adjusted to decrease overfitting and increase predictive power. The Szyszkowski equation was fitted using a Python code, yielding values for the three target variables—Cmax, K_L, and CMC—that describe the entire SFT-log(c) profile as well as their fitted confidence intervals. The ML algorithm implemented for all the datasets is the XGBoost Regressor from the scikit-learn library.

Table 2.1:Molecular descriptor libraries (22 in total), in both 2D (9) and 3D (13) used for the study

2D	3D
GSFragments (1138)	AlvaDesc (5666)
ISIDA Fragments (45)	CDK (256)
JPlogP (9)	Chemaxon (499)
MAP 4 (1024)	Dragon (S270)
Mold2 (777)	Inductive (S4)
MNA (22)	Kraken x (124)
QNPR (37)	MERA (S29)
SIRMS (161)	MERSY (42)
Structural Alerts (3256)	MOPAC2016 (35)
	MORDRED (1826)
	Pydescriptor (16251)
	RDKit (3317)
	Spectrophores (144)

The various descriptor packages are divided into 2D and 3D dimensions according to the calculated descriptors. Generally speaking, 3D descriptor libraries were better at predicting Cmax than 2D libraries. The 2D and 3D descriptor libraries both made good predictions for log(K_L) and log(CMC).

A comparative study on the use of machine learning methods for fast estimation of CO_2-brine IFT was done by Zang et al (2020). IFT data used for this study were obtained from pendant drop analysis consisting of both pure and impure CO_2. other input features used were selected based on their non-negligible effect on CO_2-brine IFT, they include pressure, temperature, nonovalent cation (Na^+ and K^+) and bivalent cation (Ca^{2+} and Mg^{2+}) molalities in the liquid phase, and methane and nitrogen fractions in CO_2. 9 machine learning techniques were implemented (GPR, SVM, KRR, MLP, CART, RF, AdaBoost, GBDT, and XGBOOST) to carryout this estimation. The table 2.2 below details the optimal settings of the hyperparameters for some of the algorithms.

Table 2.2: Optimal Settings of Hyperparameters

METHOD	HYPERPARAMETER	OPTIMUM
RF	Min. no. of samples to split a node	2
	Min no. of samples at a leaf node	1
AdaBoost	Max. tree depth	10
	Learning rate	0.2
GBDT	Max. tree depth	3
	Learning rate	0.2
XGBoost	Max. tree depth	5
	Learning rate	0.1

Three statistical indicators—the mean absolute error (MAE), mean relative error (MRE), root mean square error (RMSE), and coefficient of determination (R^2)—were used to assess and compare the performance and accuracy of the estimation models.

Okon et al (2024) developed Four models to predict crude oil brine IFT (two for each algorithm - GBDT and ADABOOST) based on varying number of dependent variables and base learners. The models were fitted with surfactants and crude oil-brine IFT data, after which they were trained, tested and evaluated to determine the best model. In the first stages of simulation, the impact and effects of training the models with varying data sizes, functional forms and decision-making processes to predict are examined. The models were then evaluated using the root mean squared error (RMSE), coefficient of determination (R^2), standard deviation (SD) and the average absolute relative deviation (AARD) statistical metrics as is the best practice for predictive ML models. The dependent variables used

for this study include surfactant critical micelle concentration, hydrophilic-lipophilic balance, molecular packing parameter, solubility ratio, density and molecular weight as well as the IFT yields for the surfactants used.

III. DESCRIPTION OF VARIABLES

As summarized in the preceding heading several frameworks have been deployed in predicting interfacial tension depending on the goals of the research. In this section insight will be provided on the choice of variables selected for each of the previously stated works with regards to their importance and effect on the target variables for the various scenarios. Generally variables for ML inputs are selected based on the level of impact they have on the target variable and their availability from experimental data (I.e. how easily they can be gotten through experimentation).

3.1.0 Effect of Crude Oil Total Acid Number (TAN) on Crude oil/brine IFT

Crude oil's Total Acid Number (TAN) indicates how acidic the oil is—more precisely, how many acidic compounds are there in the oil. The interfacial tension (IFT) between crude oil and brine can be impacted by the TAN of crude oil (Amar et al, 2019). In general, the IFT between the crude oil and brine rises in tandem with the TAN of the crude oil. (Amar et al, 2019) The reason for this is that the crude oil's acidic compounds may interact with the brine's ions, increasing the surface tension where the two fluids meet. Increased IFT between crude oil and brine may affect techniques used in enhanced oil recovery (EOR), among other oil recovery processes. (Sharma et al, 2000) The efficiency of the oil recovery process can be decreased by higher IFT, which can make it harder for the oil to pass through the reservoir rock and be replaced by injected fluids.Therefore, when developing and putting into practice oil recovery processes, it is crucial for oil producers to take the TAN of crude oil and its possible impact on IFT into account. Surfactants and other chemicals that lower surface tension and increase oil recovery efficiency can be used as strategies to lessen the impact of high TAN on IFT.

3.1.1 Effect of brine solution pH on crude oil/brine IFT

An important factor influencing the interfacial tension (IFT) between brine and crude oil is the pH of the brine solution. A solution's pH is a gauge of its acidity or alkalinity; higher pH values denote higher alkalinity and lower pH values higher acidity. As the pH of the brine solution rises, the IFT between crude oil and brine generally tends to decrease. (Farhadi et al, 2020) This is due to the fact that the pH of the solution affects the surface charge of the oil-water interface. Higher pH levels make the brine solution more alkaline, which causes negatively charged hydroxide ions to form at the interface. The crude oil's acidic constituents may interact with these ions, lowering the IFT between the two fluids.On the other hand, the brine solution becomes more acidic at lower pH levels, which may cause positively charged hydrogen ions to form at the interface. (Daaou et al, 2011) The crude oil's acidic constituents may interact with these ions, raising the IFT between the two fluids. Thus, managing the brine solution's pH can play a significant role in maximizing oil recovery procedures, including enhanced oil recovery (EOR) methods. Optimizing the pH of the brine solution to reduce the interfacial tension (IFT) between brine and crude oil can enhance the effectiveness of oil extraction and recovery from reservoirs. To improve oil recovery operations and optimize the IFT, additives or chemicals that adjust pH can also be used.

3.1.2 Effect of phase inversion temperature on crude oil/brine IFT

The Phase Inversion Temperature (PIT) is the temperature at which a change in the relative concentrations of surfactant, oil, and water phases can occur in a system. (Friberg et al, 2011) The PIT can have a significant impact on the interfacial tension (IFT) between crude oil and brine when surfactants are present in the system.(Friberg et al, 2011) When the system is above the PIT, the surfactant molecules tend to adsorb more at the oil-water interface, leading to a lower IFT between the crude oil and brine. This is because the surfactant molecules can reduce the surface tension at the interface, allowing for better mixing and interaction between the oil and water phases.The surfactant molecules may not adsorb at the interface as well

when the system is below the PIT, which would raise the IFT between the crude oil and brine. In procedures like enhanced oil recovery (EOR), where lowering the IFT is crucial for enhancing oil displacement and recovery, (Zhang et al, 2020) this could result in decreased efficiency. Thus, to maximize the efficiency of surfactant-based processes in the oil industry, it is imperative to comprehend how the phase inversion temperature affects the IFT between crude oil and brine. Operators can manipulate the IFT to optimize oil recovery operations and boost overall efficiency by varying the system temperature in relation to the PIT and the surfactant concentrations accordingly.

3.1.3 Effect of Surfactant HLB on crude oil/brine IFT

A surfactant's hydrophilic (attracting water) and lipophilic (attracting oil) properties are balanced out by something called the hydrophilic-lipophilic balance (HLB). An important factor influencing the interfacial tension (IFT) between crude oil and brine is a surfactant's HLB value. Higher HLB surfactants are generally more hydrophilic and have a tendency to be more successful in reducing the IFT between crude oil and brine. (Massaweh et al, 2020) This is because the surface tension between the oil and water phases is lowered by hydrophilic surfactants' enhanced ability to interact with water molecules at the interface.Lower HLB surfactants, on the other hand, are more lipophilic and might not be as useful in lowering the IFT between brine and crude oil. It is possible that these surfactants are more suited to the oil phase and less effective at encouraging mixing and interaction between the water and oil phases. For the purpose of maximizing the effectiveness of surfactant-based procedures used in the oil industry, such as enhanced oil recovery (EOR), it is crucial to choose surfactants with the right HLB value. (Massaweh et al, 2020) The IFT between crude oil and brine can be effectively lowered by operators by selecting surfactants with the ideal balance of hydrophilic and lipophilic properties. This increases the effectiveness of oil displacement and recovery operations.

3.1.4 Effect of surfactant CMC on crude oil/brine IFT

A surfactant's Critical Micelle Concentration (CMC) is the concentration at which its molecules group together to form micelles in a solution. In systems where surfactants are present, the interfacial tension (IFT) between crude oil and brine can be significantly influenced by the CMC. Surfactant molecules exist in the solution as individual monomers below the CMC, and their ability to reduce the IFT between crude oil and brine may be lessened. Surfactant molecules, however, begin to aggregate to form micelles as the concentration of the surfactant rises and approaches the CMC. (Massaweh et al, 2020) The surface tension and IFT between the two phases can then be lowered by these micelles adsorbing at the oil-water interface.As a result, the IFT between crude oil and brine may be significantly reduced in the presence of surfactants above the CMC. (Massaweh et al, 2020) For procedures like enhanced oil recovery (EOR), where reducing the IFT is crucial to increasing oil displacement and recovery efficiency, this may be advantageous. In the oil industry, surfactant-based process design must take the CMC of surfactants into account. Operators can effectively lower the IFT between crude oil and brine by optimizing the surfactant concentration above the CMC, which will improve oil recovery and operational performance.

3.1.5 Effect of temperature and pressure on CO_2-brine IFT

The interfacial tension (IFT) between carbon dioxide (CO_2) and brine is influenced by temperature and pressure in a CO_2-brine system. (Cheng et al, 2023) Both temperature and pressure play crucial roles in determining the IFT between CO_2 and brine in various applications, such as carbon capture and storage (CCS) and enhanced oil recovery (EOR). The IFT between CO_2 and brine typically decreases as temperature rises. (Cheng et al, 2023) The reason for this is that elevated temperatures have the potential to improve the solubility and mobility of CO_2 in the brine, resulting in an enhanced interphase interaction. However, depending on the particular makeup of the CO_2-brine system and the experimental setup, the relationship between temperature and IFT can be complicated and change.Pressure has an impact on the IFT between brine and CO_2. Generally speaking, the IFT between CO_2 and brine can drop as pressure

rises. (Xing et al, 2013) Increased pressures may encourage CO_2 to dissolve in the brine, which would improve two-phase mixing and lower the IFT. At higher pressures, the impact of pressure on the IFT between CO_2 and brine is usually more noticeable. (Xing et al, 2013) This effect can be influenced by various elements, including the composition of CO_2 and brine as well as the presence of surfactants or additives.IFT between the two phases can be greatly influenced by the interaction of temperature and pressure in a CO_2-brine system. With CO_2-related processes like CCS and EOR, where controlling the IFT between CO_2 and brine is crucial for effective operation and successful results, it is crucial to comprehend and optimize these factors during the design and implementation phases.(Xing et al, 2013)

3.1.6 Effect of nonovalent cations on CO_2/brine IFT

Nonovalent cations, such as calcium (Ca^{2+}) and magnesium (Mg^{2+}), can have a significant impact on the interfacial tension (IFT) between carbon dioxide (CO_2) and brine in CO_2-brine systems. The presence of nonovalent cations can influence the properties of the brine and affect the interactions between CO_2 and the aqueous phase, leading to changes in the IFT (Gary et al, 2022). Understanding the effects of monovalent cations on CO_2/brine IFT is crucial for enhancing carbon capture and storage (CCS) and enhanced oil recovery (EOR). Monovalent cations influence the total salinity of the brine, which in turn affects the IFT between the brine and CO_2. The presence of nonovalent cations at the CO_2-brine interface can modify the ion distribution, resulting in changes in the surface tension and IFT. Additionally, monovalent cations such as Ca^{2+} and Mg^{2+} can adsorb at the CO_2-brine interface and alter its surface characteristics, potentially affecting the interactions between the brine and CO_2. (Mutallep et al, 2018) The interactions between nonovalent cations and surfactants in the CO_2-brine system can also play a role in the interfacial behavior of the system and impact the surfactants' ability to lower the IFT between brine and CO_2. Therefore, it is important to consider the various ways in which nonovalent cations affect the IFT between brine and CO_2 for successful CCS and EOR processes.

3.1.7 Effect of Bivalent cations of CO_2/ brine IFT

Interfacial tension (IFT) between CO_2 and brine can be significantly impacted by bivalent cations, such as calcium (Ca^{2+}) and magnesium (Mg^{2+}). The system's surface tension may change as a result of these cations' interactions with the water molecules at the CO_2 and brine interface. Bivalent cations can generally reduce the interfacial tension between CO_2 and brine by interacting with the water molecules at the interface to form complexes.(Chan et al, 2024) This may result in a lower IFT due to a more stable interface and less energy needed to keep it that way. Furthermore, bivalent cations may have an impact on the interface's wettability, which may have an additional effect on the IFT between brine and CO_2. For instance, calcium ions can make the interface more wettable, which lowers the IFT. In general, lower IFT values in CO_2/brine systems may result from the presence of bivalent cations, which may have an impact on several processes including geologic carbon storage, enhanced oil recovery, and CO_2 sequestration. (Chan et al, 2024) Optimizing these processes and increasing their efficiency requires an understanding of the function of bivalent cations in these systems.

3.1.8 Effect of Langmuir Constant in IFT Reduction

The efficacy of a surfactant in decreasing surface tension in an aqueous solution is largely dependent on the Langmuir constant. (Czajka et al, 2015) The affinity of surfactant molecules for the interface between the aqueous solution and air is measured by the Langmuir constant, also referred to as the equilibrium constant. A greater affinity of the surfactant molecules for the interface is indicated by a higher Langmuir constant,(Y. Nakama, 2017) which results in a more efficient reduction of surface tension. This indicates that there is a greater chance for the surfactant molecules to adsorb to the interface and create a stable monolayer, which lowers the aqueous solution's surface tension. the effectiveness of a surfactant in reducing surface tension in water is determined by its Langmuir constant. (Czajka, 2015) A higher Langmuir constant means that the surfactant molecules are more strongly attracted to the water's surface, resulting in a greater reduction of surface tension. Conversely, a lower Langmuir constant indicates weaker attraction, resulting in a less successful reduction of surface tension and

potentially leading to a less stable monolayer. Generally, a higher Langmuir constant leads to a more effective reduction of surface tension, while a lower Langmuir constant results in a less significant decrease. It is important to note that the surfactant's ability to adsorb to the interface plays a crucial role in reducing surface tension in an aqueous solution.

IV. ALGORITHMS

4.1.0 Support Vector Regression (SVR)

Support Vector Regression (SVR) is an established and sophisticated machine learning methodology that endeavors to approximate the relationship between input variables and a continuous target variable while minimizing prediction errors. (Zarandi et al, 2020; Ghahfarokhi, 2022) Unlike Support Vector Machines (SVMs), which are predominantly utilized for classification tasks, SVR is primarily concerned with identifying a hyperplane that closely fits data points while allowing some deviation. (Danesh et al, 2022) SVR is widely acknowledged for its efficacy in handling non-linear relationships between independent and dependent variables, thanks to its use of kernel functions (such as linear, polynomial, radial basis function, sigmoid). Moreover, it is renowned for its robustness to outliers and underpinned by solid theoretical foundations. SVR relies on a subset of data points referred to as support vectors to forecast continuous outcomes such as IFT. Its utility for IFT prediction has been well-documented across varying conditions, demonstrating its versatility and practicality in this field.

4.1.1 Gradient Boosting Decision Tree (GBDT)

Gradient Boosting Decision Trees (GBDT) is a powerful ensemble machine learning algorithm that combines the principles of gradient boosting and decision trees. (Liu et al, 2022) In the gradient boosting step, the algorithm determines the residuals from the base model, which are the differences between the actual target values and the predicted values. These residuals represent the errors produced by the base model. The algorithm begins by creating a single decision tree as the foundational model. This tree is constructed by recursively dividing the data according to the feature that yields the maximum information gain at each node. The tree's prediction is the average of the target variable across the leaf nodes. To fix the mistakes made by the base model, the algorithm creates more decision trees. Each new tree is trained using the residuals from the older trees to lower prediction errors. (Liu et al, 2022) The learning rate parameter manages the contribution of each tree to the final prediction. It helps to avoid overfitting and enhance the model's performance. (Wu et al, 2021) A lower learning rate produces a more conservative model that performs better when applied to new data. The final prediction produced by the GBDT algorithm is the total of the predictions made by each individual tree, weighted by the learning rate. The algorithm iteratively creates new trees and updates the predictions until a predetermined number of trees are reached or a stopping criterion is satisfied (Rashidi-Khaniabadi et al, 2023).

4.1.2 Extreme Gradient Boosting (XGBoost)

XGBoost (Extreme Gradient Boosting) is an optimized and efficient implementation of the gradient boosting algorithm that has gained popularity in the machine learning community for its performance and scalability. (Zhang et al, 2018) With XGBoost, errors from earlier models are corrected by building a sequential ensemble of weak learners, or decision trees, in accordance with the gradient boosting principles. Through the addition of new trees that forecast the residuals of the older trees, the algorithm minimizes a loss function. To avoid overfitting, XGBoost includes L1 and L2 regularization terms in the objective function. By imposing a penalty on large coefficients, these regularization terms penalize the model's complexity and enhance the model's generalizability. XGBoost finds the best splits for every tree node using an effective algorithm. It employs an approach known as the approximate greedy algorithm to find the best split points based on the information gain at each node, as opposed to thoroughly searching every possible split. When a tree reaches maturity, XGBoost uses pruning to eliminate nodes that don't make a big difference in the model's performance. (Chen et al, 2016) Pruning keeps trees from overfitting and helps to make them less complex. During the tree-building process, XGBoost learns the optimal imputation strategy to handle missing values in the dataset. When a missing value occurs in a feature, it automatically determines which way to go

and assigns it to the left or right child node depending on the learned direction. XGBoost leverages distributed computing and parallel processing to train models more quickly on big datasets. It is engineered for efficiency and scalability. For even quicker training times, GPU acceleration is also supported.Cross-validation functionality is integrated into XGBoost to assess model performance and adjust hyperparameters. By doing so, overfitting is avoided and the ideal set of hyperparameters for the model is chosen. (Pedregosa et al, 2011; zhang et al, 2018)

4.1.3 Categorical Boosting (CATBoost)

CatBoost is a gradient boosting algorithm specifically made to function well with data that contains categorical features. One-hot encoding or label encoding are examples of pre-processing that is not necessary when using CatBoost to handle categorical features in the data. By directly handling categorical variables, it reduces overfitting and enhances model performance through an effective technique known as "ordered boosting." CatBoost employs gradient boosting, which builds an ensemble of weak learners (decision trees) in a stepwise manner to fix the mistakes made by the earlier models. (He et al, 2014) By incorporating fresh trees that forecast the residuals of the old trees, the algorithm reduces a loss function. The feature importance scores that CatBoost offers show how much each feature contributes to the model's predictions. Making educated decisions regarding feature selection and model interpretation can be facilitated by having a clear understanding of the relative importance of the various features in the dataset. L2 regularization is a feature of CatBoost that helps to improve the model's generalization and prevent overfitting. By penalizing large coefficients in the model, regularization helps limit the likelihood of overfitting and regulate the complexity of the trees (Piero et al; 2023). Several strategies, including early stopping, learning rate scheduling, and feature importance-based pruning, are included in CatBoost to prevent overfitting. These methods aid in the development of a more reliable and broadly applicable model that functions well with unknown data. A large number of hyperparameters are available in CatBoost, which can be adjusted to maximize the model's performance. In order to tune hyperparameters, the algorithm offers grid search and random search in addition to integrated cross-validation to assess model performance. CatBoost supports distributed computing and parallel processing and is built for scalability and efficiency. It is appropriate for a variety of machine learning tasks due to its ability to manage sizable datasets with millions of samples and thousands of features. (Huang et al, 2019; Dorogush et al, 2018)

4.1.4 Adaptive Boosting (ADABOOST)

A machine learning technique called AdaBoost (Adaptive Boosting) is used to combine weak classifiers into a strong classifier, hence improving their performance. Using the training data, the algorithm iteratively trains a series of weak classifiers, giving the misclassified instances greater weights with each iteration. (Debjani et al, 2020) A weighted majority vote is used to combine the predictions of all the weak classifiers to create the final prediction. same weight is given to each training instance at the beginning of the AdaBoost (Adaptive Boosting) process. The total weight is standardized to one. Every instance in the training set has an equal weight at first. The training process uses these weights to establish the relative importance of each instance. (Neha et al, 2020) Using the initial weights, the training data is used to train the first weak learner. A simple model, like a decision tree with a finite depth, is usually the weak learner. The algorithm determines the error rate—the sum of the weights of the incorrectly classified instances divided by the total weight of all instances—after training the weak learner. The instances that were incorrectly classified have heavier weights than the correctly classified instances, which have lower weights. (Sahoo et al, 2020) In the following iteration, this highlights the instances that were incorrectly classified. The updated training data is used to train the subsequent weak learner, and the weights of the instances are modified in accordance with their classification from the previous iteration. until a stopping criterion is satisfied or for a predetermined number of iterations, the final two steps are repeated. A strong learner that does well on the training set is produced when each successive weak learner concentrates more on the cases that the earlier learners misclassified. Ultimately, by giving each weak learner a weight determined by their performance, the weak learners are combined to create a strong learner. All of the

weak learners' predictions are combined to create the final forecast, with the more accurate weak learners' predictions receiving more weight. (Vikram et al, 2020)

4.1.5 Elastic Net Regression (EN)

The regularization method known as Elastic Net regression combines the penalties of Ridge (L2 norm) and Lasso (L1 norm) regression. It is especially helpful for high-dimensional datasets with multicollinearity and is used to address some of the shortcomings of these individual techniques. (Friedman et al, 2010) The Elastic Net regression model aims to minimize the following objective function:

Loss function + α * L1 penalty + β * L2 penalty (1)

Here, the loss function represents the error between the predicted values and the actual values, α and β are hyperparameters that control the strength of the L1 and L2 penalties, respectively.The L1 penalty encourages sparsity by shrinking some coefficients to zero, effectively performing feature selection. This helps in reducing the model complexity and improving interpretability.The L2 penalty penalizes large coefficients and helps in reducing the impact of multicollinearity by shrinking the coefficients towards zero. The hyperparameters alpha (α) and beta (β) play a significant role in determining the balance between L1 and L2 penalties. Larger values of alpha promote sparsity or feature selection, while larger values of beta reduce the impact of multicollinearity. Elastic Net regression model is trained by minimizing the objective function using optimization methods like gradient descent. During training, the model adjusts the coefficients of the features to find the best fit while also penalizing large coefficients to prevent overfitting. Elastic Net regression combines the advantages of Lasso and Ridge regression by conducting feature selection and handling multicollinearity simultaneously. It selects valuable features by reducing less important ones to zero (L1 penalty) and minimizing the effect of correlated features (L2 penalty). After the model is trained, it can be used to make predictions on new data. (Hastie, 2009)The coefficients learned during training are used to calculate the predicted values based on the input features.

V. RESULTS AND CONCLUSION

Zhang et al evaluated the accuracy of the estimation models using the mean absolute error (MAE), mean relative error (MRE) the root mean square error matrices (RMSE) and the coefficient of determination (R^2). The assessment matrices for the ML techniques with reference to the testing set are compiled in Table 5.0 Given its relatively high MAE/MRE/RMSE and R^2 of less than 0.9, the GPR method has the lowest accuracy of all the methods examined. A high number of data points deviate from the true values, as previously observed, which is consistent with the low accuracy GPR and further suggests that the overfitting problem is linked to the GPR.

Table 5.0: Evaluation matrices for ML methods in the testing stage

METHOD	MAE (mN/m)	MRE (%)	RMSE (mN/m)	R^2
GPR	3.05	7.87	4.26	0.881
MLP	1.62	4.48	2.22	0.967
SVM	1.60	4.16	2.24	0.967
KRR	1.67	4.41	2.38	0.963
DT	1.83	4.77	2.73	0.951
RF	1.60	4.17	2.26	0.967
ADABOOST	1.69	4.43	2.28	0.966
GBDT	1.07	2.61	1.40	0.987
XGBOOST	0.90	2.37	1.28	0.989

Regression models' robustness and generalization ability are widely acknowledged as the first important criteria to consider when evaluating their quality. It is not advisable to pursue further applications with a model (such as the GPR in this study) that can replicate the training set but is unable to predict the testing sets with acceptable accuracy. Based on this, the Xgboost outperforms the other ML models.

In the study conducted by Ali et al, the ML models developed were evaluated based on the coefficient of determination, average percent relative error (APRE, %), root mean square error (RMSE), standard deviation (SD), and average absolute percent relative error (AAPRE, %) statistical parameters. The GBRT performed better than the DT and ET algorithms as evident in the crossplots shown in figure 5.1 below:

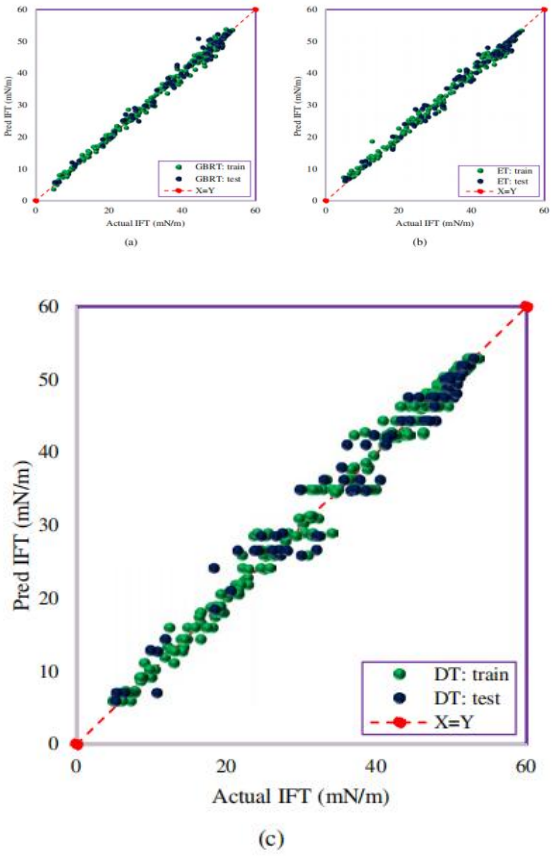

Figure 5.1: Cross plots of the developed models; (a) GBRT, (b) ET, (c) DT. (Ali et al, 2023)

The findings of Kirche's work demonstrated that all of the ML techniques used had comparable error rates. In addition to having larger errors, the linear regression (LR) approach is comparable to the machine learning techniques. Although the gain is not very great when compared to the far more practical LR model, it is clear that the ML algorithms could be used to describe the IFT database.

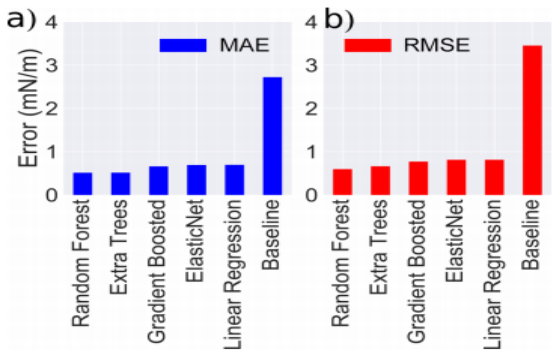

Figure 5.2: (a) Mean absolute error (MAE) and (b) root mean square error (RMSE) obtained for each estimator when applied to the test data set. These values are lower than the Naive baseline. (Kirch et al, 2020)

For the molecular dynamics simulation, Testing the simplified feature space (P, N, A, and DensityOil) using the Remesal et al. and Kakati and Sangwai data sets revealed that the average error for the former was 2% and the average error for the latter was 9%, with the larger errors for the toluene. Figure 5.3 displays a comparison of the model with data from the literature. With the exception of the toluene IFTs, which are located on the lower left of the graph, the model's transferability is evident.

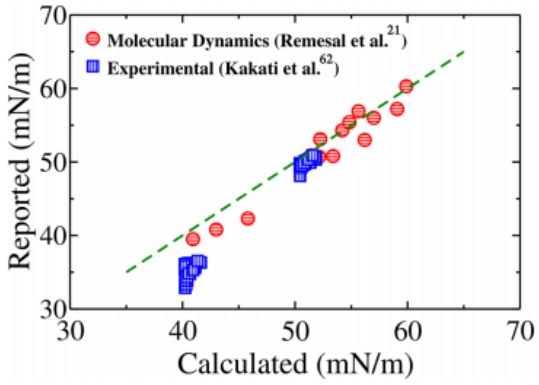

Figure 5.3: The trained model's predicted IFT values are compared to those reported by Kakati and Sangwai and Remesal et al. Experimental data from Kakati and Sangwai are represented by blue squares, and data from Remesal et al. molecular dynamics are shown in red circles. For x = y, use the dashed line.

The gradient boosted method's feature importance rank showed that the density and oil fractions were the key factors affecting the oil/brine IFT. It's interesting to note that the oil features obscured the influence of the ionic composition by predominating the IFT changes. The only other significant factor was the overall salinity of the brine, excluding the oil properties.In conclusion, one useful tool for addressing the issues facing the oil industry is the creation of consistent IFT databases from molecular simulations. Several models can be applied with well-organized available data to further predict nonsimulated scenarios with high transferability.

Expanding those databases to take into account the thermal conditions of the reservoir, a greater variety of brine compositions, and more intricate interfaces could be possible.

The results from Amar et al (2019) is summarized in the table 5.1 below. He evaluated the performance accuracy of the developed models using the RMSE, AARD, R^2 and SD statistical metrics.

Table 5.1: Statistical indexes of the established models

TESTING RESULT				
	RMSE	AARD	R^2	SD
Model 1 - Gradient Boosting Trees	1.2548	1.7933	0.9946	0.0647
Model 1 - AdaBoost Support vector Regression	1.0584	2.0829	0.9966	0.0434
Model 2 - Gradient Boosting Trees	2.0587	2.4825	0.9875	0.1067
Model 2 - AdaBoost Support vector Regression	2.0185	2.6990	0.9875	0.0817

Also from the results of the study conducted by Okon et al, the GBDT model-2 gave the best performance with an R^2 value of 99.70%, RMSE of 0.103, AARD of 1.32% and SD value of 0.0327. the table 5.2 shows the comparison of prediction of the GBDT model-2 with actual experimental values.

Table 5.2: Comparison of Prediction of the GBDT Model-2 with Actual Experimental Values

Experimental Values	GBDT Model-2 Predictions	Deviations
0.0070	0.00300	0.0011
0.0031	0.00280	0.0003
0.0026	0.00260	0.0000
0.3520	0.28000	0.0720
0.0082	0.00820	0.0000
0.0026	0.00258	0.0002
0.0350	0.03380	0.0012
1.3000	1.30000	0.0000
4.4000	4.40000	0.0000

From Table 5.2 the total deviations in the sample data is about 7.5% of the total predictions which invariablyshows an accuracy of 93% which is in line with the overall accuracy of the entire dataset prediction as shown in the R^2 result of 99.7%. figure 5.2 provides graphical insight on the prediction accuracy of the GBDT model-2 with only one significant deviation as shown.

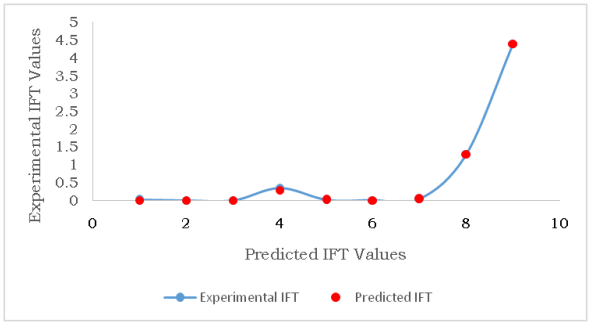

Figure 5.4: Comparison of Predicted and Actual IFT Values of the GBDT Model-2

The results obtained from the various reviewed publication shows that the boosting technique is much more effective in carrying out these predictions with higher accuracies than empirical and regular ML techniques. Figure 5.5 shows the R^2 values of some of the algorithm results discussed.

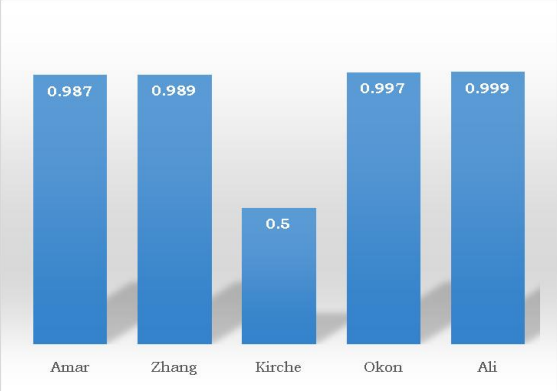

Figure 5.5: Coefficient of determination (R^2) result comparison of different authors

Conclusively, different ML frameworks have been implemented over the years to carryout prediction of several crude reservoir properties, overcoming major system complexities. In this article, the GBDT, ADAboost, CATboost and XGboost algorithms are reviewed spotlighting the different modelling frameworks that have been applied in predicting various crude oil reservoir interfacial tension property, comparing their output with regular ML techniques (non-boosting) and empirical correlations.

REFERENCES

[1] Bui, T. *et al.* Water/oil interfacial tension reduction—An interfacial entropy driven process. *JPCCP* 23(44), 25075–25085 (2021).

[2] Kalam, S., Khan, M. R., Shakeel, M., Mahmoud, M. & Abu-khamsin, S. *Smart Algorithms for Determination of Interfacial Tension (IFT) Between Injected Gas and Crude Oil-Applicable to EOR Projects* (Middle East Oil, Gas and Geosciences Show/OnePetro, 2023).

[3] Garmsiri, H. *et al.* Stability of the emulsion during the injection of anionic and cationic surfactants in the presence of various salts. *Sci. Rep.* 13(1), 11337 (2023).

[4] Shafei, M., Kazemzadeh, Y., Martyushev, D. A., Dai, Z. & Riazi, M. Efect of chemicals on the phase and viscosity behavior of water in oil emulsions. *Sci. Rep.* 13(1), 4100 (2023).

[5] Kalatehno, J. M. & Khamehchi, E. A novel packer fuid for completing HP/HT oil and gas wells. *J. Petrol. Sci. Eng.* 203, 108538 (2021).

[6] Pereira, L. M., Chapoy, A., Burgass, R. & Tohidi, B. Interfacial tension of CO2+ brine systems. *Exp. Predict. Model.* 103, 64–75 (2017).

[7] Kim, B. *et al.* Ensemble machine learning-based approach for predicting of FRP-concrete interfacial bonding. *Mathematics* 10(2), 231 (2022).

[8] Tadros, T. Gibbs adsorption isotherm. In *Encyclopedia of Colloid and Interface Science* (Tadros, T. ed.). 626 (Springer, 2013). Sibanda, D., Oyinbo, S. T. & Jen, T.-C. A review of atomic layer deposition modelling and simulation methodologies: Density functional theory and molecular dynamics. *Nanotechnol. Rev.* 11(1), 1332–1363 (2022).

[9] Singh, S. K., Chaurasia, A. & Verma, A. Basics of Density Functional Teory, Molecular Dynamics, and Monte Carlo Simulation Techniques in Materials Science. In *Coating Materials: Computational Aspects, Applications and Challenges* (eds Verma, A. *et al.*) 111–124 (Springer, 2023).

[10] Zhao, X., Duan, W., Zeng, X. & Liu, Y. J. Measurements of surface tension of R1234yf and R1234ze (E). *Int. J. Refrig.* 63(1), 21–26 (2018). Clegg, C. *Contact Angle Made Easy: Carl Clegg* (2013). DA. *Standard Test Methods for Surface and Interfacial Tension of Solutions of Paints, Solvents, Solutions of Surface-Active Agents, and Related Materials. Annual Book of ASTM Standards*. (American Society for Testing and Materials, 2014).Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. J. Comput. Phys. 1995, 117, 1−19.

[11] Brown, W. M.; Wang, P.; Plimpton, S. J.; Tharrington, A. N. Implementing Molecular Dynamics on Hybrid High Performance Computers−Short Range Forces. Comput. Phys. Commun. 2011, 182, 898−911.

[12] Brown, W. M.; Kohlmeyer, A.; Plimpton, S. J.; Tharrington, A. N. Implementing Molecular Dynamics on Hybrid High Performance

Computers−Particle−Particle Particle-Mesh. Comput. Phys. Commun. 2012, 183, 449−459.

[13] Brown, W. M.; Yamada, M. Implementing Molecular Dynamics on Hybrid High Performance Computers-Three-Body Potentials. Comput. Phys. Commun. 2013, 184, 2785−2793.

[14] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg. Scikit-Learn: Machine Learning in Python. Journal of Machine Learning Research 2011, 12, 2825−2830.toward predicting interfacial tension of impure and pure co2-brine systems using robust correlative approaches chen wei-yu,lin sun, jiyong zhou, xuguang li*, liping huang, guang xiamengand kui wang

[15] Gary R. Jerauld, Aboulghasem Kazemi,An improved simple correlation for accurate estimation of CO2-Brine interfacial tension at reservoir conditions,Journal of Petroleum Science and Engineering,Volume 208, Part C,2022,109537,ISSN 0920-4105,https://doi.org/10.1016/j.petrol.2021.109537.

[16] Mutallep, Mhriay, Jiang, Lanlan, Fu, Jin, Wang, Zhe, Yu, Tianbo, Lu, Zilong, Liu, Yu; 2018/06/01; Effects of Na+, K+, Ca2+, and Mg2+ cations on CO2–brine interfacial tension under offshore storage conditions, Vol 8; DO - 10.1002/ghg.1787; JO - Greenhouse Gases: Science and TechnologyER -

[17] Cheng Zhang, Milei Wang,-CO2/brine interfacial tension for geological CO2 storage: A systematic review,Journal of Petroleum Science and Engineering,Volume 220, Part A,2023,111154,ISSN 0920-4105,https://doi.org/10.1016/j.petrol.2022.111154.(https://www.sciencedirect.com/science/article/pii/S0920410522010063)

[18] Wanli Xing, Yongchen Song, Yi Zhang, Masahiro Nishio, Yangchun Zhan, Weiwei Jian, Yong Shen, Research Progress of the Interfacial Tension in Supercritical CO2-water/oil System,Energy Procedia,Volume 37, 2013,Pages 6928-6935,ISSN 1876-6102,https://doi.org/10.1016/j.egypro.2013.06.625.(https://www.sciencedirect.com/science/article/pii/S1876610213008680)

[19] Osama Massarweh, Ahmad S. Abushaikha, The use of surfactants in enhanced oil recovery: A review of recent advances, Energy Reports, Volume 6,2020, Pages 3150-3178, ISSN 2352-4847, https://doi.org/10.1016/j.egyr.2020.11.009. (https://www.sciencedirect.com/science/article/pii/S2352484720314347)

[20] Friberg, Stig & Corkery, Robert & Blute, Irena. (2011). Phase Inversion Temperature (PIT) Emulsification Process. Journal of Chemical & Engineering Data. 56. 4282–4290. 10.1021/je101179s.

[21] Zhang, Madhar Sahib Azad, J.J. Trivedi, IFT or wettability alteration: What is more important for oil recovery in oil-wet formation?, Fuel, Volume 291, 2021,119986,ISSN 0016-2361, https://doi.org/10.1016/j.fuel.2020.119986. (https://www.sciencedirect.com/science/article/pii/S0016236120329823)

[22] Hamed Farhadi, Shahab Ayatollahi, Mobeen Fatemi, The effect of brine salinity and oil components on dynamic IFT behavior of oil-brine during low salinity water flooding: Diffusion coefficient, EDL establishment time, and IFT reduction rate, Journal of Petroleum Science and Engineering, Volume 196,2021, 107862, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2020.107862.(https://www.sciencedirect.com/science/article/pii/S0920410520309220)

[23] Mortada Daaou, Dalila Bendedouch,Water pH and surfactant addition effects on the stability of an Algerian crude oil emulsion,Journal of Saudi Chemical Society, Volume 16, Issue 3, 2012, Pages 333-337,ISSN 1319-6103, https://doi.org/10.1016/j.jscs.2011.05.015.(https://www.sciencedirect.com/science/article/pii/S1319610311001141)

[24] Sharma, Mukul & Filoco, P.. (2000). Effect of Brine Salinity and Crude-Oil Properties on Oil Recovery and Residual Saturations. SPE Journal - SPE J. 5. 293-300. 10.2118/65402-PA.

[25] Menad Nait Amar, Mohammadhadi Shateri, Abdolhossein Hemmati-Sarapardeh, Alireza

Alamatsaz, Modeling oil-brine interfacial tension at high pressure and high salinity conditions, Journal of Petroleum Science and Engineering, Volume 183, 2019, 106413, ISSN 0920-4105, https://doi.org/10.1016/j.petrol.2019.106413. (https://www.sciencedirect.com/science/article/pii/S0920410519308344)

[26] Dargahi-Zarandi, A., Hemmati-Sarapardeh, A., Shateri, M., Menad, N. A. & Ahmadi, M. Modeling minimum miscibility pressure of pure/impure CO2-crude oil systems using adaptive boosting support vector regression: Application to gas injection processes. *J. Petrol. Sci. Eng.* 184, 106499 (2020).

[27] Ghahfarokhi, A. J. & Amar, M. N. Well production forecast in Volve feld: Application of rigorous machine learning techniques and metaheuristic algorithm. *J. Petrol. Sci. Eng.* 208, 109468 (2022).

[28] Talebkeikhah, M. *et al.* Experimental measurement and compositional modeling of crude oil viscosity at reservoir conditions. *J. Taiwan Inst. Chem. Eng.* 109, 35–50 (2020).

[29] Nait Amar, M. & Zeraibi, N. A combined support vector regression with frefy algorithm for prediction of bottom hole pressure. *SN Appl. Sci.* 2(1), 23 (2020).

[30] Amar, M. N., Zeraibi, N. & Jahanbani, G. A. Applying hybrid support vector regression and genetic algorithm to water alternating CO2 gas EOR. *Greenh. Gases Sci. Technol.* 10(3), 613–630 (2020). Sethi, A. Support vector regression tutorial for machine learning. *Stat. Comput.* 14, 1–5 (2020).

[31] Zamani, M. G. *et al.* A multi-model data fusion methodology for reservoir water quality based on machine learning algorithms and bayesian maximum entropy. *J. Clean. Prod.* 416, 137885 (2023).

[32] Danesh, A., Ehsani, M., Moghadas Nejad, F. & Zakeri, H. Prediction model of crash severity in imbalanced dataset using data leveling methods and metaheuristic optimization algorithms. *Int. J. Crashworthin.* 27(6), 1869–1882 (2022).

[33] Wanan Liu, Hong Fan, Meng Xia, Credit scoring based on tree-enhanced gradient boosting decision trees, Expert Systems with Applications, Volume 189, 2022, 116034, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2021.116034. (https://www.sciencedirect.com/science/article/pii/S0957417421013798)

[34] Rashidi-Khaniabadi, A., Rashidi-Khaniabadi, E., Amiri-Ramsheh, B., Mohammadi, M.-R. & Hemmati-Sarapardeh, A. Modeling interfacial tension of surfactant-hydrocarbon systems using robust tree-based machine learning algorithms. *Sci. Rep.* 13(1), 10836 (2023).

[35] Wu, J., Li, Y. & Ma, Y. (eds.) Comparison of XGBoost and the neural network model on the class-balanced datasets. In *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer* (ICFTIC) (IEEE, 2021). Chen, T. & Guestrin, C. (eds.) Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016).

[36] Larestani, A., Mousavi, S. P., Hadavimoghaddam, F. & Hemmati-Sarapardeh, A. Predicting formation damage of oil felds due to mineral scaling during water-fooding operations: Gradient boosting decision tree and cascade-forward back-propagation network. *J. Petrol. Sci. Eng.* 208, 109315 (2022).

[37] Peiro Ahmady Langeroudy, K., Kharazi Esfahani, P. & Khorsand Movaghar, M. R. Enhanced intelligent approach for determination of crude oil viscosity at reservoir conditions. *Sci. Rep.* 13(1), 1666 (2023).

[38] Dorogush, A.V., Ershov, V. & Gulin, A. CatBoost: Gradient boosting with categorical features support. *arXiv Preprint* arXiv:181011363 (2018).

[39] Huang, G. *et al.* Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. *J. Hydrol.* 574, 1029–1041 (2019).

[40] X. He, J. Pan, O. Jin, T. Xu, B. Liu, T. Xu, Y. Shi, A. Atallah, R. Herbrich, S. Bowers, et al. Practical lessons from predicting clicks on ads at facebook. In Proceedings of the Eighth

International Workshop on Data Mining for Online Advertising, pages 1–9. ACM, 2014

[41] Debjani Chakraborty, Ahona Ghosh, Sriparna Saha, Chapter 2 - A survey on Internet-of-Thing applications using electroencephalogram, Editor(s): Valentina Emilia Balas, Vijender Kumar Solanki, Raghvendra Kumar,Emergence of Pharmaceutical Industry Growth with Industrial IoT Approach,Academic Press,2020,Pages 21-47,ISBN 9780128195932,https://doi.org/10.1016/B978-0-12-819593-2.00002-9. (https://www.sciencedirect.com/science/article/pii/B9780128195932000029)

[42] Neha Sharma, Usha Batra, Sherin Zafar, Chapter 1 - A neoteric swarm intelligence stationed IOT–IWD algorithm for revolutionizing pharmaceutical industry leading to digital health,https://doi.org/10.1016/B978-0-12-819593-2.00001-7. (https://www.sciencedirect.com/science/article/pii/B9780128195932000017)

[43] Subhashree Sahoo, Amiya Bhusan Bagjadab, Sushree Bibhuprada B. Priyadarshini, Chapter 10 - Smart pills: a complete revolutionary technology than endoscopy,Emergence of Pharmaceutical Industry Growth with Industrial IoT Approach,Academic Press,2020,Pages 275-303,ISBN 9780128195932, https://doi.org/10.1016/B978-0-12-819593-2.00010-8. (https://www.sciencedirect.com/science/article/pii/B9780128195932000108)

[44] Vikram Puri, Raghvendra Kumar, Dac Nhuong Le, Sandeep Singh Jagdev, Nidhi Sachdeva,https://doi.org/10.1016/B978-0-12-819593-2.00011-X. (https://www.sciencedirect.com/science/article/pii/B978012819593200011X)

[45] *Friedman, Jerome; Trevor Hastie; Rob Tibshirani (2010).* "Regularization Paths for Generalized Linear Models via Coordinate Descent". Journal of Statistical Software. *33 (1): 1–22.* doi*:*10.18637/jss.v033.i01. PMC 2929880. PMID 2080872

[46] Hastie, T.; Tibshirani, R.; Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction; Springer Science & Business Media, 2009.

[47] Okon, John and Udoh, Tinuola and Emenka, Blessed, Prediction of Crude Oil-Brine Interfacial Tension Based on Surfactant Characteristics Using Artificial Intelligence. Available at SSRN: https://ssrn.com/abstract=4770150 or http://dx.doi.org/10.2139/ssrn.4770150