



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Prediction of Crude Oil-Water Interfacial Tension with Surfactants and Nanomaterials Using Machine Learning

Nathaly Salomé Garzón Ramos

Facultad de Minas
Universidad Nacional de Colombia
Medellín, Colombia
2025

Prediction of Crude Oil-Water Interfacial Tension with Surfactants and Nanomaterials Using Machine Learning

Nathaly Salomé Garzón Ramos

Master's thesis submitted as a partial requirement for the degree of:

Master's Degree in Engineering - Analytics

Advisors

John Willian Branch Bedoya, Ph.D

Camilo Andrés Franco Ariza, Ph.D

Facultad de Minas

Universidad Nacional de Colombia

Medellín, Colombia

2025

To my family for their strength, motivation, and incredible support; to my friends for their company and understanding; and to my travel companion for being an inspiration and holding my hand throughout this investigation.

Acknowledgments

To professors John Willian Branch and Camilo Andrés Franco for their supervision, guidance, and support throughout the development of this master's thesis. Additionally, I extend my thanks to the Surface Phenomena Research Group – Michael Polanyi and the Artificial Intelligence Research and Development Group (GIDIA) for their valuable feedback and comments, which were remarkable in making this project possible.

This master's study was made possible thanks to funding from the Francisco José de Caldas Fund, MINCIENCIAS and the National Hydrocarbons Agency (ANH) through contract No. 112721-282-2023 (Project 1118-1035-9300) with the Universidad Nacional de Colombia - Sede Medellín and PAREX RESOURCES COLOMBIA AG SUCURSAL.

Predicción de la Tensión Interfacial Crudo-Agua con Surfactantes y Nanomateriales empleando Aprendizaje de Máquinas

Resumen

La predicción precisa de la tensión interfacial (IFT) es un factor crítico para el diseño y la optimización de los procesos de recuperación mejorada de petróleo por medios químicos (cEOR). Este estudio se centra en la aplicación de cuatro modelos predictivos (RF, ET, GBRT, XGBoost) para la IFT en sistemas con surfactantes y nanomateriales. Para ello, se utilizó un conjunto de datos experimentales y basados en la bibliografía con 551 puntos de datos, caracterizado por una distribución desbalanceada compuesta por un 75 % de mediciones de IFT inferiores a $20,55 \text{ mN}\cdot\text{m}^{-1}$ (sistemas con aditivos) y un 25 % de experimentos de control con valores más altos. Esta estructura de datos se conservó intencionadamente para garantizar la representatividad fenomenológica del modelo. El rendimiento del modelo se evaluó utilizando métricas de rendimiento (R^2 , RMSE, MAE), gráficos residuales y curvas de aprendizaje. El análisis de las curvas de aprendizaje reveló que el rendimiento del modelo deja de mejorar después de aproximadamente 200 muestras de entrenamiento, lo que demuestra que incorporar datos adicionales similares no es beneficioso. Los resultados confirman que el modelo de bosque aleatorio es la herramienta más robusta para predecir la IFT con un R^2 del 85 % y subrayan que una composición de datos representativa es más crucial que un equilibrio estadístico estricto, lo que ofrece una valiosa orientación para optimizar los futuros esfuerzos de recopilación de datos.

Palabras Clave: Aprendizaje de Máquinas, Tensión Interfacial, Surfactante, Nanofluidos, Recobro Mejorado de Petróleo.

Abstract

Accurate prediction of interfacial tension (IFT) is a critical factor for the design and optimization of chemical enhanced oil recovery (cEOR) processes. This study focuses on the application of four predictive models (RF, ET, GBRT, XGBoost) for IFT in systems with surfactants and nanomaterials. For this purpose, an experimental and literature-based dataset with 551 datapoints was used, characterized by an imbalanced distribution composed of 75% IFT measurements below $20.55 \text{ mN}\cdot\text{m}^{-1}$ (systems with additives) and 25% from control experiments with higher values. This data structure was intentionally preserved to ensure the model's phenomenological representativeness. Model performance was evaluated using performance metrics (R^2 , RMSE, MAE), residual plots, and learning curves. Analysis of the learning curves revealed that the model's performance stops improving after approximately 200 training samples, demonstrating that incorporating similar additional data is not beneficial. The results confirm that the random forest model is the most robust tool for predicting IFT with an R^2 of 85% and underscore that a representative data composition is more crucial than a strict statistical balance, offering valuable guidance for optimizing future data collection efforts.

Keywords: Machine Learning, Interfacial Tension, Surfactant Nanofluids, Enhanced Oil Recovery.

Contents

	Pág.
Figure List.....	XIII
Table List	XIV
Abbreviations	XV
1. Introduction	17
1.1 Motivation	17
1.2 Problem statement.....	19
1.3 Objectives.....	20
1.3.1 General Objective.....	20
1.3.2 Specific Objectives.....	20
1.4 Thesis Scope.....	21
1.5 Document Structure	21
2. Theoretical framework	22
2.1 Machine Learning Models for Predicting the Interfacial Tension of Nanofluids containing Surfactants.....	22
2.1.1 Bagging Learning.....	24
2.1.2 Boosting Learning.....	25
2.2 Phenomena Related to the Interfacial Tension of Nanofluids containing Surfactants.....	26
3. Systematic literature review.....	30
3.1.1 Impact of Physicochemical Parameters on the Conformation of Surfactant-Modified Nanofluids.....	32
3.1.2 Implementation of Machine Learning Models to Predict the Physical Properties of Surfactant-Modified Nanofluids.	35
4. Methodology for Predicting Crude Oil-Water Interfacial Tension Using Surfactants and Nanomaterials with Machine Learning	38
4.1 Variable Selection in Interfacial Tension Prediction using Machine Learning	39
4.2 Interfacial Tension Data Compilation: Experimental and Literature Sources	41
4.2.1 Experimental Data Collection on Interfacial Tension using Surfactants and Nanomaterials	41
4.2.2 Literature Data Collection on Interfacial Tension Measurements using Surfactants and Nanomaterials	42
4.3 Machine Learning Models Selection for Interfacial Tension Prediction in Nanofluids containing Surfactants.....	43
5. Results	45

5.1	Interfacial Tension Measurements of Nanofluids with Surfactants	45
5.2	Analysis and Description of the Dataset's Structure	47
5.3	Performance of the Implemented Machine Learning Models.....	49
6.	Conclusions and Future Work.....	56
7.	References.....	58

Figure List

	Pag.
<i>Figure 1-1. Colombia's Total Energy Supply for 2023. Adapted from [2].</i>	17
<i>Figure 1-2. Evolution of Energy Supply in Colombia since 2000. Adapted from [2].</i>	18
<i>Figure 2-1. Classification of Machine Learning Approaches. Adapted from [28].</i>	23
<i>Figure 2-2. Ensemble Learning: Bagging & Boosting techniques. Adapted from [30].</i>	24
<i>Figure 3-1. Histogram of yearly scientific output based on sections A1 and A2 of the systematic literature review.</i>	32
<i>Figure 4-1. Graphical representation of oil droplet displacement through a pore throat driven by reduced interfacial tension.</i>	40
<i>Figure 5-1. IFT vs Concentration CQD.</i>	46
<i>Figure 5-2. IFT vs Concentration CQD in the presence of Surfactant.</i>	47
<i>Figure 5-3. Dataset distribution</i>	48
<i>Figure 5-4. Spearman correlation matrix for the continuous variables.</i>	49
<i>Figure 5-5. Performance metrics.</i>	50
<i>Figure 5-6. Evaluation of Bagging Models: Actual vs. Predicted Values and Residual Plots for (A) Random Forest and (B) Extra Trees</i>	51
<i>Figure 5-7. Evaluation of Boosting Models: Actual vs. Predicted Values and Residual Plots for (A) Gradient Boost Regressor Tree and (B) Extreme Gradient Boost</i>	52
<i>Figure 5-8. Learning Curves for Bagging and Boosting Models using RMSE: (A) Extra Trees and Random Forest, and (B) Gradient Boosting and XGBoost.</i>	54

Table List

	Pag.
<i>Table 3-1 Research equations used for the systematic review of the literature.....</i>	31
<i>Table 3-2. Predictive Models and Performance Metrics in IFT Prediction Related Studies.</i>	36
<i>Table 4-1. Hyperparameters used to train the models.....</i>	44
<i>Table 5-1. Descriptive Statistics of Interfacial Tension in the Dataset</i>	48
<i>Table 5-2. Tuned Model Hyperparameters.....</i>	55

Abbreviations

Abbreviation	Term
<i>EOR</i>	Enhanced Oil Recovery
<i>cEOR</i>	Chemical Enhanced Oil Recovery
<i>OPEX</i>	Operational Expenditures
<i>OOIP</i>	Original Oil in Place
<i>CQD</i>	Carbon Quantum Dots
<i>AI</i>	Artificial Intelligence
<i>ML</i>	Machine Learning
<i>IFT</i>	Interfacial Tension
<i>CMC</i>	Critical Micelle Concentration
<i>RF</i>	Random Forest
<i>ET</i>	Extra Trees
<i>GBRT</i>	Gradient Boost Regressor Tree
<i>XGBoost</i>	Extreme Gradient Boost
R^2	Coefficient of Determination
<i>RMSE</i>	Root Mean Square Error
<i>MAE</i>	Mean Absolute Error

1. Introduction

1.1 Motivation

Colombia's energy matrix, as illustrated in *Figure 1-1*, is predominantly composed of hydrocarbons, [1] which make up 75.2% of the total energy supply as of 2023. [2] While some energy sources are used directly, others are converted into fuels or electricity for final consumption. [3] Moreover, fuel demand has steadily increased over the past decade as confirmed by the International Energy Agency (IEA) and shown in *Figure 1-2*. This positive trend is expected to persist, emphasizing the energy sector's crucial role in the nation's sustainable development and the importance of ensuring a consistent and reliable energy supply.

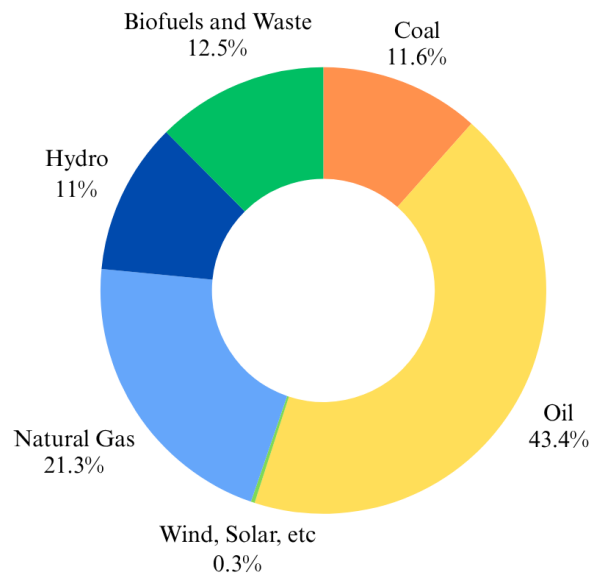


Figure 1-1. Colombia's Total Energy Supply for 2023. Adapted from [2].

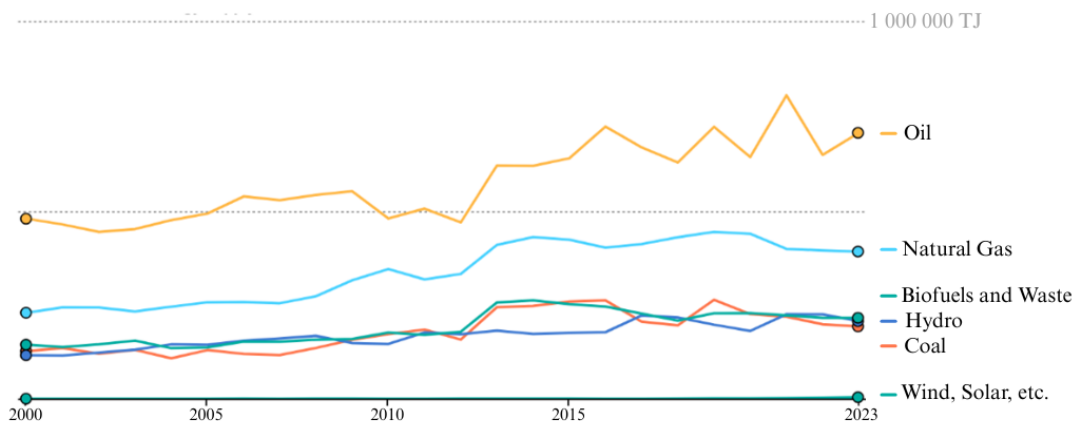


Figure 1-2. Evolution of Energy Supply in Colombia since 2000. Adapted from [2].

The success of modern enhanced oil recovery (EOR) operations, which can elevate the final recovery factor from a conventional ceiling of 40% to as high as 65% of the original oil in place (OOIP), [4] is fundamentally dependent on the precision of their chemical design. Among these advanced techniques, chemical EOR (cEOR) through surfactant injection has emerged as a particularly promising field, [5] reflected in a global market valued at \$900 million in 2021 and projected to grow. [6] However, the efficacy of this method is contingent upon solving a complex, multi-variable optimization problem: designing a chemical formulation that is perfectly adapted to the unique geology and fluid characteristics of a given reservoir.

The core challenge lies in identifying a formulation that simultaneously satisfies multiple, often conflicting, criteria. It must generate ultralow interfacial tension (IFT) to mobilize residual oil, exhibit minimal adsorption onto the rock matrix, maintain stability under harsh reservoir conditions of high temperature and pressure, and provide effective mobility control, all while remaining economically viable. Given that chemical costs can constitute up to 30% of a project's operational expenditures (OPEX), the selection of a suboptimal formulation is a primary cause of both technical and financial failure. [7, 8] This complexity is further amplified by recent innovations incorporating nanomaterials into surfactant systems. While the addition of nanoparticles like Carbon Quantum Dots (CQDs) [9] can significantly boost the oil recovery factor by an additional 14% to 24% [10-13] by mitigating surfactant retention [14, 15] and improving sweep efficiency, [16, 17] it also adds another layer of variables to the optimization process.

Traditionally, the development of these formulations has relied on resource-intensive, trial-and-error laboratory experiments, creating a significant bottleneck in project timelines and budgets. To overcome this, the oil and gas industry is increasingly turning to data-driven methodologies. [18-20] Machine learning (ML), in particular, offers a powerful alternative, capable of rapidly and accurately predicting the performance of chemical systems from experimental data. With a proven track record of optimizing well performance and analyzing geological data, it can help to identify potential reservoirs and select optimal recovery techniques. [21] Su et al. [22] used this method to reveal the distribution and relationships between different wells and fluid parameters, which helped in determining the most effective recovery technique. The ability of ML algorithms to discern complex, non-linear relationships between formulation components and reservoir conditions to forecast key performance indicators like IFT, establishes the technology as a powerful tool for enhancing efficiency, increasing production, and reducing operating costs.

This technological imperative holds particular relevance for Colombia's energy sector. As of 2024, only 25% of the country's oil production used EOR methods, [23] highlighting a vast, untapped potential to enhance recovery factors and bolster national reserves. The development of a predictive tool to design more effective and cost-efficient cEOR formulations is, therefore, not merely a technical exercise but a strategic necessity. To address this need, this research will implement machine learning models capable of mapping these intricate non-linear relationships. By leveraging data from an experimental scenario and existing literature, the analytical focus will be on evaluating the synergistic effects of key variables on the resulting interfacial tension (IFT) in complex surfactant-nanomaterial systems.

1.2 Problem statement

The central analytical challenge in optimizing chemical enhanced oil recovery is a high-dimensional, non-linear prediction problem: forecasting complex and sensitive properties such as interfacial tension (IFT). [24] The input space for this problem is defined by the characteristics and concentrations of formulation components (e.g., surfactants,

nanomaterials, oil, and water) and the specific physicochemical conditions of a reservoir. [25-27] The current method for navigating this space relies on a trial-and-error experimental paradigm. This approach constitutes a dispersed, expensive, and analytically inefficient method of data acquisition to capture the complex, multi-variable relationships governing overall performance. This research directly addresses this predictive gap by proposing the implementation of machine learning models to approximate the complex function linking chemical inputs and reservoir conditions to IFT. By leveraging available experimental and literature data, this approach reframes the challenge with a varied dataset, with the goal of deploying a fast and reliable predictive model to improve the performance of costly and time-consuming laboratory tests. The implementation of this analytical tool would directly mitigate financial risk by identifying high-performing formulations at an early stage, accelerate the innovation and deployment of novel chemical solutions, and ultimately improve the economic viability of EOR projects.

1.3 Objectives

1.3.1 General Objective

Propose a model for the prediction of crude oil-water interfacial tension with surfactants and nanomaterials using machine learning.

1.3.2 Specific Objectives

- Identify the relevant variables for the predictive model to estimate the crude-water interfacial tension in the presence of surfactants and nanomaterials using machine learning.
- Construct a data set that combines experimental data and data collected from literature on crude-water interfacial tension in the presence of surfactants and nanomaterials.
- Select the predictive model to determine the oil-water interfacial tension with surfactants and nanomaterials using machine learning.
- Evaluate the performance of the selected model using performance metrics.

1.4 Thesis Scope

The scope of this thesis covers two core components: an experimental section and the subsequent application of predictive machine learning models. The experimental phase involves the measurement of interfacial tension (IFT) in a crude oil-brine system from a Colombian field, assessing surfactant formulations with four distinct carbon quantum dots (CQDs). This analysis is complemented by the implementation and comparative analysis of four supervised machine learning regression models to identify the most robust method for predicting IFT from a dataset characterized by significant challenges, including low data volume, high imbalance, and inherent phenomenological variability. Finally, this study serves as an initial framework for integrating machine learning as a complementary tool in the screening process for EOR projects, demonstrating the potential to extrapolate experimental results and accelerate the design of cEOR formulations.

1.5 Document Structure

This thesis is structured into six chapters to fulfill the objectives outlined in the first chapter. The second chapter introduces a theoretical framework for the machine learning models used in this research, along with a discussion of interfacial tension phenomena involving surfactants and nanomaterials.

The third chapter provides a systematic review of the literature on interfacial tension measurements using surfactants and nanomaterials, as well as the application of machine learning models to predict interfacial tension. In the fourth chapter, the methodology used in this research is detailed, including the construction of the dataset from both experimental measurements and literature sources, a descriptive analysis of the dataset, and the implementation of models along with error metrics.

The fifth chapter presents the results obtained, while the sixth chapter summarizes the conclusions, future directions, and opportunities arising from this work.

2. Theoretical framework

This chapter outlines the fundamental concepts of the machine learning models applied to the subject of study. Additionally, it discusses the phenomenological aspects of the interactions between nanomaterials and surfactants with interfacial tension. This ensures that the results produced by the machine learning models align with the observed phenomena.

2.1 Machine Learning Models for Predicting the Interfacial Tension of Nanofluids containing Surfactants

Machine Learning (ML) models have emerged as powerful tools for predicting interfacial tension (IFT) in complex systems, such as nanofluids containing surfactants. ML algorithms can directly learn intricate non-linear relationships from experimental data, without requiring an explicit mechanistic understanding of every underlying physicochemical phenomenon. [28] This capability is particularly advantageous for such systems, where the effects of nanoparticles on surfactant performance can be highly non-intuitive.

In this context, the main objective is to implement these predictive tools that can rapidly and accurately estimate IFT under diverse conditions, thereby significantly reducing experimental burden, optimizing formulation design, and accelerating process development. To achieve this goal, different algorithmic approaches can be applied, including but not limited to supervised, unsupervised, and ensemble learning algorithms, as shown in *Figure 2-1*.

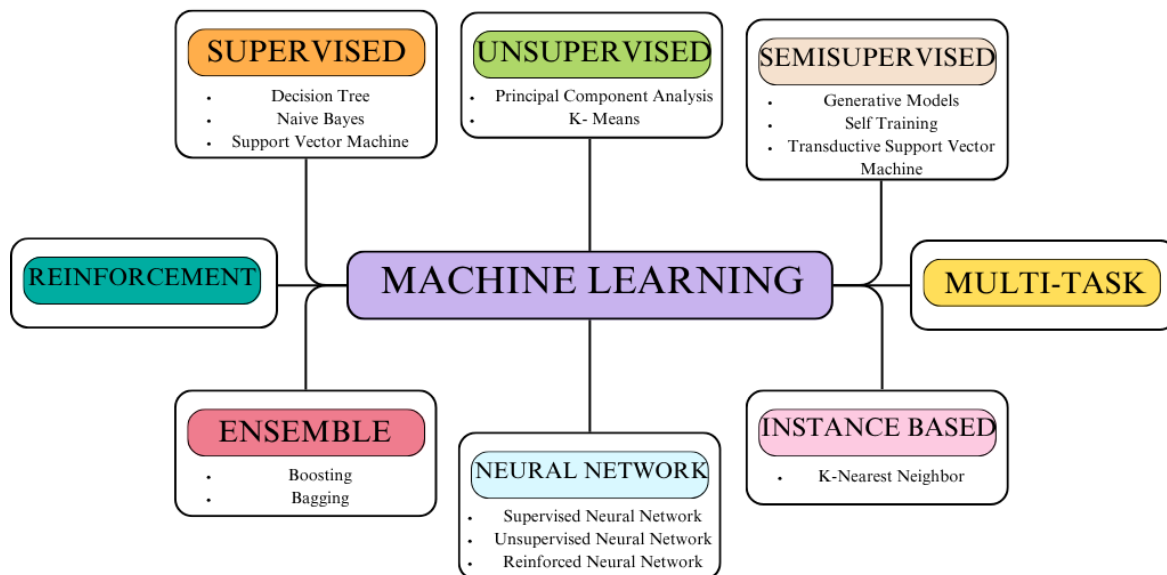


Figure 2-1. Classification of Machine Learning Approaches. Adapted from [28].

The complex and multivariate nature of interfacial tension in systems containing surfactants and nanomaterials makes its prediction a challenging task suited for supervised machine learning. The objective is to predict a continuous numerical output (IFT) based on a set of known input variables. In machine learning, this type of task is formally defined as a regression problem.

Supervised learning involves training an algorithm with a labeled dataset, where both the input variables and the desired output variable are known. [28] This prior knowledge allows the model to learn the underlying patterns and relationships within the data. To leverage the predictive power of supervised learning while addressing the nonlinearity and variability present in IFT data, ensemble modeling strategies are particularly effective [29] as illustrated in *Figure 2-2*. By employing these methods, it is possible to explore different learning techniques to achieve the most accurate predictions of IFT, thereby effectively capturing the intricate relationships between component properties and system conditions.

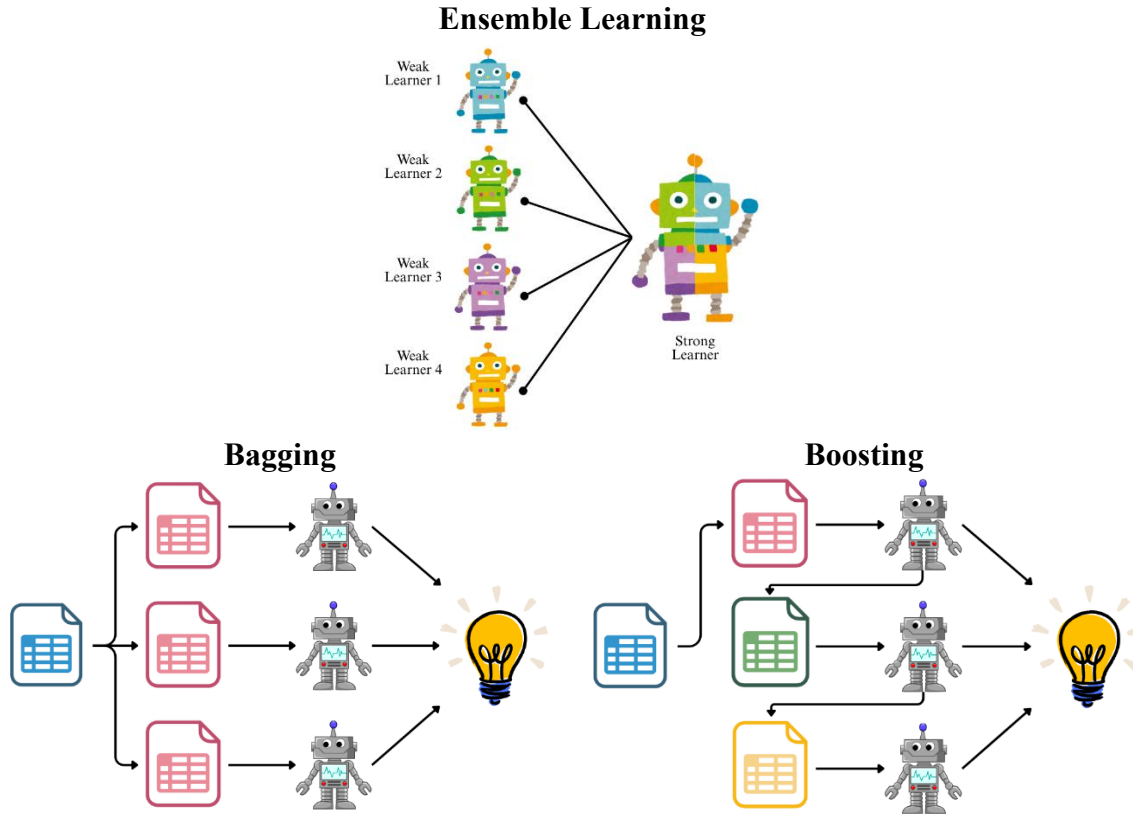


Figure 2-2. Ensemble Learning: Bagging & Boosting techniques. Adapted from [30].

2.1.1 Bagging Learning

Bagging, or Bootstrap Aggregating, involves constructing multiple decision trees independently and simultaneously. Each tree is trained on a distinct subset of the original dataset, which is created by resampling the data with replacement, a technique known as bootstrapping. The final prediction is obtained by averaging the predictions from all the individual trees. This method is highly effective for reducing model variance and minimizing overfitting, resulting in more stable predictions. [28]

2.1.1.1 Random Forest

Random forest (RF) is recognized for its ability to improve the accuracy and predictive stability of models while mitigating the tendency toward overfitting inherent in individual decision trees. [31] The operation of RF is based on two key principles that introduce randomness and, consequently, reduce the correlation between trees. First, the bagging method ensures that each tree is trained on a unique subset of the original data, created

through sampling with replacement. By averaging the predictions of these multiple trees, a significant reduction in the model's variance is achieved. Second, feature randomization introduces additional diversity: at each node in the tree, instead of considering all available features for a split, RF randomly selects a subset of these. This step is crucial for decreasing the correlation between trees, as each is trained with different views of the features, strengthening the diversity of the ensemble and its generalization ability. [32]

2.1.1.2 Extra Trees

Extra trees (ET) incorporates a strong randomization into two key aspects of the tree construction process: attribute selection and the determination of cut-off points for node division. The predictions from the multiple trees in the ensemble are aggregated to obtain the final result. [33] The primary advantages of ET lie in its ability to achieve strong variance reduction and minimize bias through explicit randomization and averaging of ensemble predictions. Additionally, ET is distinguished by its computational efficiency, consistently outperforming other tree-based methods, thanks to a simpler node splitting procedure that does not require pre-sorting of the sample. [33]

2.1.2 Boosting Learning

The boosting technique employs an iterative strategy where each new tree added to the ensemble is trained to correct the residual errors generated by the cumulative model up to that point. The process begins with a global prediction designed to minimize a predefined loss function. In each subsequent iteration, the algorithm calculates the "pseudo-residuals", which are the negative gradients of the loss function. These pseudo-residuals indicate the direction and magnitude in which the model needs to improve. A new regression tree is then trained to fit these pseudo-residuals, effectively dividing the feature space and estimating a constant value in each region to correct the errors. The final prediction of the model is the aggregate sum of the contributions from all the trees built sequentially. [34]

2.1.2.1 Gradient Boost Regressor Tree

Gradient boosted regression trees (GBRT) are used for predicting continuous numerical values. The effectiveness of a GBRT model relies heavily on the careful tuning of its key hyperparameters. From the perspective of the bias-variance trade-off, GBRT primarily aims to reduce model bias. While individual trees may exhibit high bias, their iterative combination and correction of residual errors enable GBRT to create a model with significantly lower bias than any single robust predictor. [35] Its iterative error correction process and inherent capacity to manage nonlinear and complex relationships make it a preferred choice for high-precision predictive modeling in challenging fields. [34]

2.1.2.2 Extreme Gradient Boosting

Extreme gradient boosting (XGBoost) is a notable ensemble method for using a second-order Taylor expansion to approximate the loss function. This approach enables the algorithm to capture more complex and nonlinear relationships in the data. Additionally, XGBoost incorporates a regularization term in its objective function, which penalizes model complexity and prevents overfitting through the use of alpha and lambda parameters. [36] Furthermore, XGBoost employs column sampling, similar to random forest, by randomly selecting a subset of features at each split. This technique reduces the correlation between trees and enhances the robustness of the ensemble. Lastly, XGBoost's optimized design allows for parallel and distributed computing, which accelerates training on large datasets, and it also includes features such as early stopping to further mitigate overfitting. [34]

2.2 Phenomena Related to the Interfacial Tension of Nanofluids containing Surfactants

The interfacial tension (IFT) between crude oil and water is an important physical property that significantly impacts the oil and gas industry, particularly in enhanced oil recovery (EOR) applications. EOR techniques are essential for maximizing hydrocarbon recovery by improving sweep efficiency and displacing residual oil that primary or secondary methods cannot recover. [24] At the pore level, this trapped oil is held by capillary forces, which are directly influenced by the IFT between crude oil and water. The effectiveness of any EOR

process in overcoming these capillary forces is quantitatively measured by the capillary number (N_c), a key dimensionless parameter that balances the viscous forces driving the fluid against the capillary forces that retain it. The capillary number, expressed as $N_c = \frac{v\mu}{\sigma}$ where v is flow velocity, μ is the viscosity of the injected fluid, and σ is the IFT, must be increased to a certain threshold to mobilize the trapped oil. [37]

Traditional EOR methods, particularly chemical flooding, aim to significantly reduce IFT. One key technique involves the injection of surfactants, which are amphiphilic molecules that lower IFT by adsorbing at the oil-water interface. [24] The conventional goal has been to achieve ultra-low IFT (ULIFT). [38, 39] While this approach is theoretically effective in a controlled laboratory setting, it presents considerable operational and economic challenges at a field scale. The success of this method in the lab, which creates a thermodynamically stable Winsor Type III microemulsion, contributes to more serious problems in the field. [40] This stable emulsion is difficult and costly to separate during post-production, leading to a significant increase in costs. [41] Additionally, emulsified water can cause equipment corrosion and damage refinery catalysts. These issues are not limited to surface operations. Emulsion formation can also occur within the reservoir, causing formation damage, reducing long-term permeability, and complicating future production. [42]

In addition to these operational challenges, the economic feasibility of surfactant flooding is negatively impacted by chemical retention. The high cost of surfactants is intensified by their significant loss due to adsorption onto rock surfaces. [43] The high specific surface area of reservoir rocks aggravates this issue, requiring the injection of a considerably larger quantity of surfactant to achieve the desired results.

To address these limitations, nanotechnology has emerged as a promising alternative, offering a fundamentally different approach. Nanoparticles, due to their small size and high specific surface area, can interact with reservoir fluids and rocks in innovative ways. [44] Their size allows them to easily enter the narrow pore throats of unconventional reservoirs, a key advantage over conventional chemicals that may lead to pore blockage. A significant conceptual shift introduced by nanotechnology is the transition from the singular pursuit of

ULIFT to a multi-objective optimization strategy. This practical approach aims to balance maximum oil recovery with operational efficiency and cost-effectiveness. This strategy allows for a substantial increase in the capillary number without inducing high-viscosity, stable emulsions that complicate fluid handling, thus optimizing the balance between displacement efficiency and practical viability. [45]

The true value of nanotechnology in EOR lies not in replacing surfactants but in developing a strong synergistic relationship with them. They can alter the micelle structure of surfactant mixtures, resulting in more compact and stable micelles that have a greater capacity for IFT reduction. [46] Studies have shown that nanoparticles with a similar charge to the surfactant can enhance the surfactant's surface activity, while oppositely charged nanoparticles can actively transport surfactant molecules to the interface through Brownian motion, further improving IFT reduction. [17, 47]

This synergistic interaction is particularly effective in addressing surfactant retention, which is a significant economic and performance challenge. [43] Nanoparticles provide an effective solution through a competitive adsorption mechanism. When introduced into a system with surfactants, the surfactant molecules tend to preferentially adsorb onto the surfaces of the nanoparticles, resulting in surfactant-coated nanoparticles. These coated nanoparticles serve as protective carriers, shielding the valuable surfactant molecules from premature adsorption onto the rock surface. [48] They then migrate to the oil-water interface, where a competitive adsorption process causes the surfactant to desorb from the nanoparticle and redistribute at the interface to reduce IFT. [49] This process ensures that the chemical reaches its intended target, optimizing its use and significantly reducing operational costs.

The optimal performance of these nanofluid systems relies on a complex interplay of different factors rather than a single component. For instance, a change in one factor, such as salinity, can trigger synergistic or antagonistic effects on other components. Environmental conditions, particularly salinity and temperature, play a crucial role. Salinity affects the arrangement of surfactant molecules at the interface and can lead to precipitation or aggregation of nanoparticles if the levels are excessive. [50] Additionally, temperature influences surfactant solubility and fluid viscosity, and it can enhance the transport of agents to the interface. [51] The composition of the crude oil itself is also critical, as natural polar

components like asphaltenes and resins can compete with injected agents for space at the oil-water interface. [52] Therefore, for a nanofluid system to be successful, the choice of surfactant, whether anionic, cationic, or non-ionic, and the type of nanoparticle used, such as silica, carbon-based, or polymer-coated, must be carefully considered. [53, 54] Furthermore, the concentration of these components is also important. The surfactant concentration must be above its critical micelle concentration (CMC) [55] to ensure the interface is saturated. Likewise, nanoparticle concentration must be controlled to prevent agglomeration, which can reduce effective surface area and cause pore blockage, damaging permeability and potentially increasing IFT. [56]

The complex interactions between components, concentrations, and system conditions underscore the challenge of accurately estimating interfacial tension. Therefore, it is crucial to consider the synergistic and antagonistic effects of these variables when evaluating and predicting interfacial tension behavior. The future of nano-EOR goes beyond materials development; it also requires the intelligent design and deployment of these systems. The traditional trial-and-error method is being replaced by a data-driven, predictive approach made possible by artificial intelligence (AI) and machine learning (ML). These computational tools analyze datasets from nanomaterial characterization to predict intricate interactions with reservoir fluids and rocks.

This method is based on the concept of technological convergence, which refers to the merging of previously distinct technologies to create a more powerful and comprehensive solution. [57] In this context, the main convergence occurs between nanotechnology and information technology. Nanotechnology supplies engineered materials, such as nanoparticles and surfactants, designed to modify interfacial tension. On the other hand, information technology employs AI and ML to provide an analytical framework for modeling and predicting the behavior of these materials within intricate fluid systems. This convergence establishes a predictive framework that reduces the need for extensive physical experimentation, thereby accelerating the development of highly effective and sustainable EOR solutions. [58]

3. Systematic literature review

To establish a clear relationship between the key aspects of the reference framework, a systematic review of the literature is conducted using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology. [59] PRISMA outlines essential criteria for the transparent conduct and reporting of systematic reviews. This methodology promotes methodological rigor, reproducibility, and traceability at every stage by defining clear steps for identifying, selecting, evaluating, and synthesizing relevant studies. By adhering to these well-defined procedures, it's possible to conduct a structured search for high-quality information, which not only enhances the validity of the results but also ensures that the evidence included is both relevant and reliable. This approach effectively guides the research direction from the very beginning.

First, the study period is established, covering the last decade from 2015 to 2025, to identify relevant reference articles. This timeframe allows the review to focus on the most current research in the field, taking into account the latest trends and technological advancements. Next, search equations are formulated, as shown in *Table 3-1*, to narrow down the articles of interest in databases such as Scopus and Google Scholar. The creation of these equations is essential for ensuring that the search is comprehensive and accurate by using a combination of keywords and boolean operators to capture all pertinent studies. Finally, a selection is made from the most relevant articles for the project by applying predefined inclusion and exclusion criteria, which helps to ensure the quality and relevance of the information gathered.

Table 3-1 Research equations used for the systematic review of the literature.

Section	Search Equation	Available Articles	Selected Articles
A1	1. TITLE-ABS-KEY (("EOR" OR "Enhanced Oil Recovery") AND ("Surfactant" OR "Nanoparticles") AND "Machine Learning") AND PUBYEAR > 2013 AND PUBYEAR < 2025	64	10
	2. TITLE-ABS-KEY ("Quantum dots" AND "Machine Learning" AND "Enhanced Oil Recovery") AND PUBYEAR > 2015 AND PUBYEAR < 2025		
	3. TITLE-ABS-KEY ((Enhanced Oil recovery) AND (Machine Learning) AND (IFT)) AND PUBYEAR > 2015 AND PUBYEAR < 2025		
A2	4. TITLE-ABS-KEY ("IFT" AND "Surfactant" AND ("Quantum dots" OR "Nanoparticles")) AND PUBYEAR > 2015 AND PUBYEAR < 2025 AND (EXCLUDE (EXACTKEYWORD , "Carbon Dioxide") OR EXCLUDE (EXACTKEYWORD , "Emulsification") OR EXCLUDE (EXACTKEYWORD , "Emulsion") OR EXCLUDE (EXACTKEYWORD , "Emulsions") OR EXCLUDE (EXACTKEYWORD , "Foam") OR EXCLUDE (EXACTKEYWORD , "Foam Stability") OR EXCLUDE (EXACTKEYWORD , "Microemulsions")) AND (LIMIT-TO (DOCTYPE , "ar"))	134	25

The equations in *Table 3-1* were formulated based on the proposed problem, leading to the definition of two research questions as follows:

1. What are the effects of physicochemical parameters on the interfacial tension of nanofluids containing surfactants?
2. Which machine learning models have been used to predict the interfacial tension of nanofluids containing surfactants, and what metrics were used for their evaluation?

In response to these research questions, a systematic literature review was conducted in two main sections:

- A1: Implementation of machine learning models to predict the interfacial tension of nanofluids containing surfactants.

- A2: Review of the reference framework concerning interfacial tension data in nanofluids containing surfactants.

It is important to note that the number of articles available in section A1 is low, indicating a lack of frequent publication in this area. This highlights the novelty and research potential in contrast to section A2, which shows a larger and steadily increasing number of publications over the years. This trend, as illustrated in *Figure 3-1*, suggests a greater availability of data for constructing a dataset, along with a variety of variables and prediction models useful for research. *Figure 3-1* presents a histogram of overall annual scientific production related to our keywords in the equations, showing a rising interest in recent years, although the field is still considered to be developing. This review aims to identify relevant variables for constructing the data set and to highlight promising models for implementation in the current study.

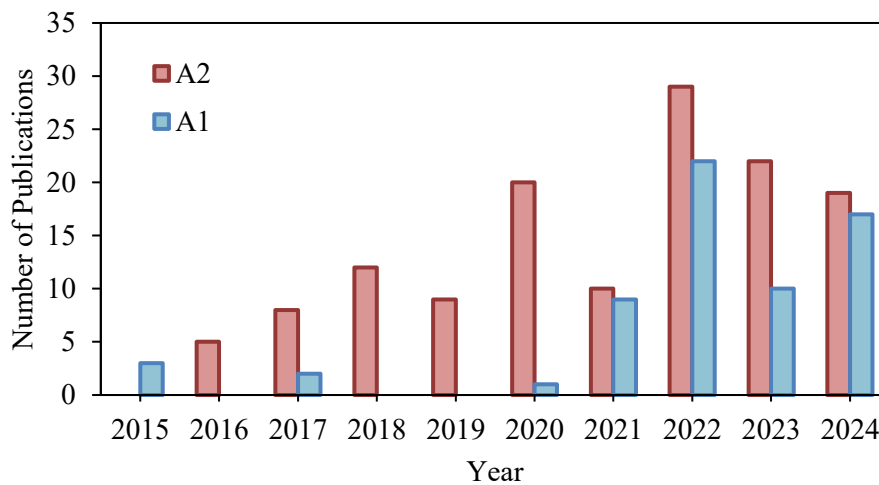


Figure 3-1. Histogram of yearly scientific output based on sections A1 and A2 of the systematic literature review.

3.1.1 Impact of Physicochemical Parameters on the Conformation of Surfactant-Modified Nanofluids

The stability, conformation, and interfacial performance of surfactant-modified nanofluids are closely related to the physicochemical parameters of the surrounding medium. Among these factors, the salinity of the aqueous medium is one of the most studied due to its impact

on both surfactants and the stability of nanoparticles. The literature indicates that controlling the ionic concentration and composition is fundamental to designing an effective formulation for reducing IFT. [17, 55] For ionic surfactants, the presence of electrolytes in the water creates an electrostatic screening effect around the charged heads of the surfactants. This effect neutralizes their mutual repulsion and enables a more compact and efficient arrangement at the oil-water interface. This denser molecular packing is a direct mechanism responsible for achieving ultra-low IFT. [50, 53] However, the literature also documents the effects of high or uncontrolled salinity, which can compromise the integrity of the entire system. A high salt concentration can lead to the destabilization and aggregation of nanoparticles. Liu et al. [47] investigated the stability of silica nanoparticles and found that while aggregation at 500 mM NaCl was partially reversible, a higher concentration of 1000 mM NaCl caused strong and irreversible aggregation. On the other hand, ionic surfactants are susceptible to the "salting-out" phenomenon, where excessive ionic strength reduces their solubility to the point of precipitation, which eliminates their effect on IFT. [55]

In addition to ionic strength, temperature also affects the interfacial properties of surfactant and nanoparticle systems. Higher temperatures can be advantageous by increasing the energy of the system. This enhanced thermal energy promotes more vigorous Brownian motion, which can accelerate the diffusion and migration of both surfactant molecules and nanoparticles to the oil-water interface. As a result, this may lead to a quicker reduction in IFT. [60] Additionally, the sensitivity analysis conducted by Rashidi-Khaniabadi et al. [61] confirmed that temperature is a significant parameter affecting IFT, ranking fourth in impact after surfactant-specific properties such as the hydrophilic-lipophilic balance (HLB).

Literature also provides evidence of thermal limitations and complex behaviors that underscore the importance of temperature control. For non-ionic surfactants, a significant limitation is the presence of a critical temperature. Above this temperature, the surfactant's hydrogen bonds with water are disrupted, leading to a loss of solubility, phase separation, and a decrease in effectiveness. [55] Other studies indicate that IFT increases with temperature. For instance, Akhlaghi et al. [50] found that for the non-ionic surfactant Triton

X-100, IFT consistently rise as the temperature increased from 20°C to 60°C. A similar trend was observed by Ivanova et al. [60] for both anionic and cationic surfactants in the absence of salt.

The interaction between temperature and other components of the system, such as salinity and nanoparticles, adds additional complexity. Ivanova et al. [60] observed a V-shaped interfacial tension pattern in surfactant-brine systems, where the IFT initially decreased to a minimum at 35°C before rising again at higher temperatures. The presence of nanoparticles can result in distinct thermal responses. For instance, Wei and Babadagli [51] observed that for a SiO₂ nanofluid, the IFT value reached a minimum as the temperature increased. The diverse and occasionally conflicting findings indicate that the impact of temperature on IFT depends on the surfactant, the brine, and the nanoparticles involved.

Finally, in addition to the properties of the medium, the concentrations of the active components are among the most fundamental parameters for achieving ultra-low IFT. The effectiveness of surfactants depends on operating at or above the CMC. [55, 62] For this reason, experimental designs often employ surfactant concentrations above the determined CMC to ensure maximum interfacial activity, such as the 1% surfactant concentration used by Nwani et al. [53] or the 2000 mg L⁻¹ used by Betancur et al. [63]. The synergy in binary surfactant systems can further lower this required threshold, as Gazem et al. [62] demonstrated by determining a CMC of 2005 ppm for an SDS-TX100 mixture, which is lower than the values of the individual surfactants.

The introduction of nanoparticles adds another layer of complexity. Zallaghi et al. (2018) tested silica nanoparticle concentrations between 500 and 5000 ppm, identifying 2000 ppm as the optimal level for maximizing incremental oil recovery during their core floods. Similarly, Mohajeri et al. (2019) found that while adding 200 ppm of SiO₂ nanoparticles to a surfactant solution improved oil recovery, higher concentrations could negatively impact the recovery due to nanoparticle agglomeration and removal of free surfactant from the bulk phase. [64, 65] This highlights that increasing nanoparticle dosage does not guarantee better results; rather, an optimal loading must be identified. [63, 66]

Consequently, recent studies have shifted focus from individual concentrations to the synergistic optimization of the entire surfactant-nanoparticle-brine system. [56, 62, 63, 67] Gazem et al. [62] reported that adding 2 wt% of SiO₂ nanoparticles decreased the CMC of their binary surfactant mixture by nearly 25%. The sensitivity analysis by Rashidi-Khaniabadi et al. [61] highlighted the interaction, identifying surfactant concentration as one of the most influential parameters on IFT.

This literature review highlights that the interfacial performance of surfactant-modified nanofluids is influenced by the complex and synergistic relationships among multiple physicochemical factors. These include the surfactant type, the nanoparticles, their respective concentrations, and the conditions of the liquid phases, such as brine salinity, crude oil type, and temperature. The non-linear and interdependent nature of these relationships makes it challenging to predict the interfacial tension using traditional analytical models. Addressing this predictive challenge paves the way for the use of machine learning models to predict the IFT of surfactant-modified nanofluids.

3.1.2 Implementation of Machine Learning Models to Predict the Physical Properties of Surfactant-Modified Nanofluids.

In recent years, machine learning (ML) has emerged as an innovative tool in petroleum engineering for modeling complex systems, accurately predicting key parameters, and optimizing processes. [29] Rashidi-Khaniabadi et al. [61] successfully implemented the gradient boost regressor tree (GBRT) along with other models to predict the IFT of surfactant-hydrocarbon systems, achieving a high coefficient of determination (R^2) of 0.9939. Their research also identified critical input parameters, such as surfactant concentration and hydrophilic-lipophilic balance. Yousefmarzi et al. [68] also used tree-based algorithms along with other models like support vector regressor (SVR) to predict oil-water IFT with an accuracy of $R^2 = 0.999$. Building on these examples related to predicting properties like interfacial tension (IFT), the application of machine learning has expanded to tackle the complexities of enhanced oil recovery processes. This development goes beyond modeling simple oil-water systems and addresses more sophisticated challenges.

These include predicting the rheological behavior of oil-surfactant-brine microemulsions [69], forecasting the oil recovery resulting from surfactant treatments [70], and identifying the key drivers that influence recovery mechanisms, such as surfactant-polymer flooding. [29]

The application of these models has also been expanded to include systems involving nanoparticles. El-Amin et al. [35] used artificial data to predict the transport of nanoparticles in porous media with two-phase flow, employing machine learning models based on decision trees and artificial neural networks. Similarly, Alvim et al. [71] investigated the effects of SiO₂ nanoparticles in enhanced oil recovery processes and discussed the potential benefits of chemical additives. These research efforts are significant because they show that it is possible to model intricate systems that include nanomaterials.

In this context, *Table 3-2* presents an overview of research conducted in the field, particularly concerning the prediction of interfacial tension. It includes information about the systems used for evaluation, the variables employed to train the models, the size of the datasets, and the models themselves, along with their respective evaluation metrics. For clarity, the best-performing model for each study is highlighted in bold, and the reported performance metrics correspond specifically to that model. The findings from these studies demonstrate the feasibility of this project, indicating that accurate predictions can be achieved even with relatively small datasets.

Table 3-2. Predictive Models and Performance Metrics in IFT Prediction Related Studies.

System Evaluated	Variables	Dataset size	Models	Metrics	Ref
H ₂ - Brine	Temperature Pressure Density	64	ENN, GPR , BT, LR	MAPE: 0.051 MAE: 0.002	[72]
Oil - Brine	Salinity Temperature	7017	GLM, GAM, RF , SVM, XGBoost, BRT,	R ² : 0.99, RMSE: 0.2, MSE: 0.02, MAE: 0.13	[73]

H ₂ – Brine + Inert gas	Temperat ure Pressure Salinity IFT, H2 CH4	2250	RF, Random Tree, LSBoost	R ² : 0.99, RMSE: 0.42, MAE: 0.30	[74]
Surfactan t - Hydrocar bons	Temperat ure Surfactan t concentra tion Alkane MW HLB, PIT	390	DT, ET, GBRT	R ² : 0.99, RMSE: 1,126	[61]
N ₂ /CO ₂ + n- Alkanes	Temperat ure Pressure Carbon number Mole fraction of N2	268	RBF , ANFIS, LSSVM, RF, MLP, ET	R ² : 0.99, RMSE: 0.115	[75]
Oil - Brine	Type of Salt Salt concentra tion, API, Temperat ure Pressure	796	CNN, AdaBoost , DT , RF, KNN, Ensemble learning, SVM, MLP- ANN	R ² : 0.97, MSE: 0.0005, MAE: 0.01	[76]
Oil - Gas Oil - Water	GOR, Oil FVF, Oil and Gas density, Gas FVF	11075	SVR , RF, DT, GB, CatBoost, XGBoost	R ² : 0.99, RMSE: 0,05, MAE: 0.03	[68]
N ₂ - Brine	Temperat ure Pressure Salinity Molality, Molarity	364	ANFIS , ANN	R ² : 0.99, MSE: 0.24, RMSE: 0.49	[77]

The systematic review of the literature indicates that, to our knowledge, while machine learning models have proven effective, most studies focus primarily on systems that incorporate only surfactants. There is limited experimental data available for more complex systems that involve the synergistic effects of nanoparticles. This highlights the need to expand and consolidate datasets for more accurate and generalizable modeling. In this context, this study aims to fill this gap by implementing predictive models trained with both experimental and bibliographic data. This approach will allow for a more precise analysis of the influence of nanomaterials and surfactants on interfacial tension, using the predictive power demonstrated in previous studies to navigate the complexities of surfactant-modified nanofluids.

4. Methodology for Predicting Crude Oil-Water Interfacial Tension Using Surfactants and Nanomaterials with Machine Learning

To address the research gap identified in the systematic literature review, this chapter outlines the methodology developed to predict the interfacial tension using machine learning models in complex surfactant-modified nanofluid systems. The process is organized into three key stages. The first stage, described in Section 4.1, discusses the selection of the input variables that overall govern interfacial tension. Next, Section 4.2 involves compiling the dataset by integrating values from both original experimental studies and existing literature. Finally, Section 4.3 details the implementation of machine learning models to determine the most effective predictive model for this application.

4.1 Variable Selection in Interfacial Tension Prediction using Machine Learning

The selection of input variables is a foundational stage in constructing a predictive model, as it defines the information the model will use to learn and, consequently, determines its capacity for generalization and accuracy. In this study, the choice of variables was guided by a rigorous methodological approach based on two criteria: their phenomenological importance, validated through the theoretical framework, and their statistical relevance, as demonstrated in previous machine learning modeling studies.

The phenomenological justification is grounded in the direct and proven influence that each variable has on the physicochemical mechanisms governing interfacial tension (IFT). Surfactants, for example, serve as the primary agents in reducing IFT, and their effectiveness critically depends on their properties. As amphiphilic molecules, surfactants migrate to the oil-water interface. As their concentration increases and approaches the critical micelle concentration (CMC), the saturation of the interface reduces the system's free energy, leading to a significant and non-linear decrease in IFT. [24]

Equally important is the hydrophilic-lipophilic balance (HLB), which quantifies a surfactant's affinity for each phase. The maximum reduction in IFT occurs only when there is an optimal HLB that balances this affinity for a specific crude oil and brine system, making it a crucial design parameter. Additionally, nanomaterials function as synergistic agents that modify the interface. Their type and concentration determine their ability to adsorb at the interface, creating a film structure and acting as carriers that enhance the local concentration of the surfactant, thereby improving its effectiveness. [46]

These interactions are influenced by the thermodynamic and chemical environment of the system. Brine salinity is particularly important, as ions in the brine neutralize the electrostatic repulsion between surfactant molecules at the interface. This neutralization allows for denser packing of the surfactants, resulting in a greater reduction in interfacial tension (IFT) until an optimal salinity is reached. [50] Additionally, the crude oil's °API

serves as an important indicator of the type of crude. This type affects how the lipophilic tail of the surfactant interacts with the oil, making it a crucial factor for the system's effectiveness. [52] Temperature, as a key thermodynamic parameter, also impacts surfactant solubility, fluid viscosity, and the kinetics of adsorption at the interface, ultimately influencing the overall energy of the system. [51]

To support this physicochemical analysis, a review of predictive modeling literature confirms that these same variables: surfactant and its concentration, hydrophilic-lipophilic balance, nanomaterial type and concentration, salt, salinity, crude oil API°, and temperature, represented in *Figure 4-1* are the most frequently and successfully used to train machine learning models, underscoring their high predictive power. Consequently, the selection strategy is grounded not only in the physical principles underlying the phenomenon but also aligns with best practices and empirical findings within the field of data analytics.

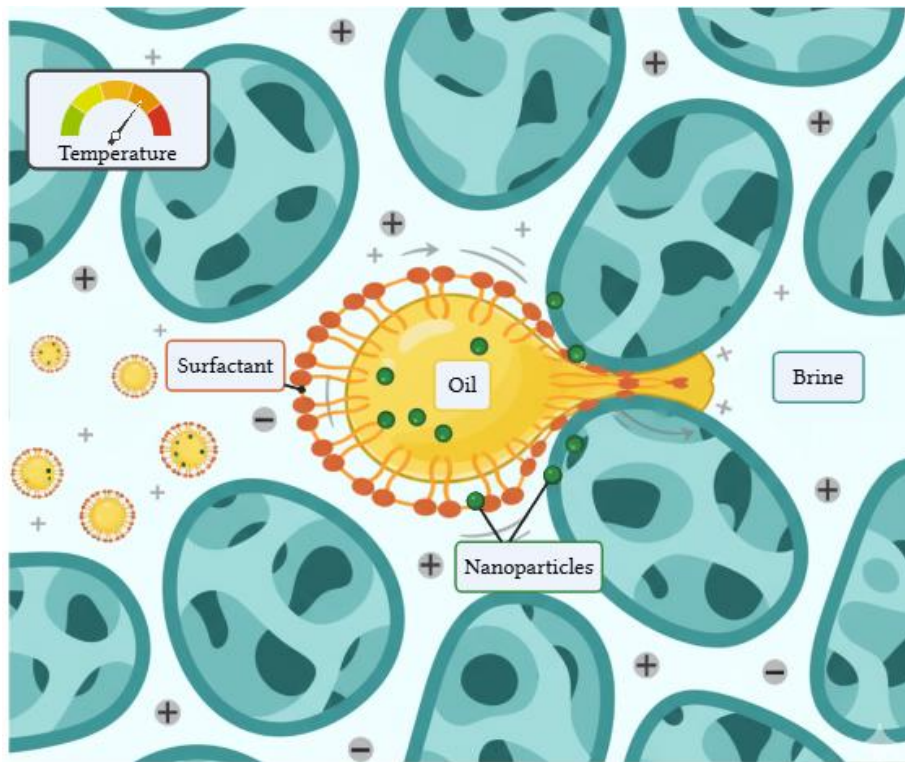


Figure 4-1. Graphical representation of oil droplet displacement through a pore throat driven by reduced interfacial tension.

4.2 Interfacial Tension Data Compilation: Experimental and Literature Sources

The second step in successfully applying predictive models is to construct a comprehensive and diverse dataset. To address the lack of data in systems that combine surfactants and nanomaterials, we adopted a dual strategy that integrates two complementary sources of information. This approach combines the generation of original experimental data, which offers precise measurements, with a compilation of data from existing literature. The synergy created by these two sources allows for the development of a dataset that is both novel and specific, while also being broad and varied. This significantly enhances the model's ability to generalize and capture a wider range of physicochemical behaviors. The following sections will detail the procedures followed for each of these data collection stages.

4.2.1 Experimental Data Collection on Interfacial Tension using Surfactants and Nanomaterials

Following the outlined methodology, the first phase involved generating the experimental data. This was achieved by measuring the interfacial tension in a system consisting of light crude oil (37.1° API) and brine, both representative of a Colombian oilfield. The brine had a pH of 8.32, a conductivity of 1727 mS·cm⁻¹, and a salinity of 917 mg·L⁻¹. The experiment evaluated the synergy of a commercial surfactant and four different carbon quantum dots (CQDs), a class of carbon-based nanomaterials: two commercial variants, designated CQDblue and CQDred, with average particle sizes of 46.7 and 84.5 nm, respectively. The other two were synthesized from molasses and coffee waste, following the hydrothermal synthesis protocol established by Rosales et al. [78], which resulted in particle sizes ranging from 15 to 100 nm.

For the experimental development, 100 mL aqueous solutions were prepared using the brine. The nanomaterial concentrations were set at 0, 50, 100, 200, and 500 mg·L⁻¹ for each nanomaterial in individual samples, while surfactant concentrations were set at either 0 or

1000 mg·L⁻¹. To ensure a homogeneous particle dispersion, each solution underwent one hour of magnetic stirring, followed by three hours of sonication.

For the measurement of interfacial tension, two complementary techniques were employed at reservoir temperature (70°C). For tensions above 1 mN·m⁻¹, an Attention Sigma 702 tensiometer based on the Du Noüy ring method was used, positioning a platinum ring at the crude-brine interface and performing automatic measurements in duplicate. For ultra-low IFT values (below 1 mN·m⁻¹), a M6500 Spinning Drop tensiometer was used. In this method, a crude oil droplet was introduced into a capillary filled with the aqueous solution, and its elongation during rotation was monitored for 1.5 hours. The IFT was determined once the system reached equilibrium, indicated by a stabilized droplet shape with no further changes in elongation.

This experimental procedure produced an initial dataset containing 34 interfacial tension measurements. Each data point connects input variables, including the nanomaterial, its concentration, the presence of a surfactant, the properties of brine and crude oil, as well as the experimental temperature, to the interfacial tension as the output variable to be predicted.

4.2.2 Literature Data Collection on Interfacial Tension Measurements using Surfactants and Nanomaterials

The second phase of constructing the dataset involved compiling interfacial tension measurements from the existing literature. A systematic review identified numerous articles that reported IFT values. To ensure the quality and relevance of the data, inclusion criteria were established to select the articles that would form the dataset. The main criteria were as follows:

- The article had to be fully accessible for review.
- It must have described all the selected input variables for model training, thus ensuring no incomplete data records were included.
- The study could present control experiments and systematically vary at least one of the variables of interest.

From these criteria, 517 data points were obtained from fourteen articles. By combining these literature-based entries with the original experimental measurements, the final

dataset of 551 data points was compiled into a tabular file format. As all records were complete, no entries were discarded. The dataset was subsequently loaded and preprocessed using the Python programming language and its open-source libraries: Pandas, Numpy, Scikit-learn, and Seaborn. To handle categorical variables, they were numerically encoded using the OneHotEncoder technique from Scikit-learn to incorporate their effect into the model's training and prediction.

It is important to note that due to the inclusion of control experiments, the dataset exhibits an imbalanced distribution. This condition is shown by a significant difference in magnitude, where 75% of the data shows IFT values below $20.55 \text{ mN}\cdot\text{m}^{-1}$, demonstrating the effect of surfactants and nanomaterials, while the remaining 25% corresponds to IFT values between 20.55 and $36.36 \text{ mN}\cdot\text{m}^{-1}$ from systems without additives. Despite this imbalance, all extreme values were retained in the analysis without modification to ensure phenomenological representativeness and to guarantee the model is trained on the full range of possible outcomes.

In the context of interfacial tension, Spearman's correlation is beneficial, as it enables the identification of relationships that do not necessarily follow linear patterns between the variables being analyzed. The Spearman correlation gives a coefficient that expresses the strength and direction of the relationship between two phenomena, with +1 being a direct correlation, 0 no correlation, and -1 an inverse correlation. [79]

4.3 Machine Learning Models Selection for Interfacial Tension Prediction in Nanofluids containing Surfactants.

To predict IFT values based on the proposed conditions, different learning models were implemented. The predictive models were random forest (RF), extra trees (ET), gradient boost regressor tree (GBRT), and extreme gradient boost (XGBoost), used from the sklearn library. Each of the implemented models underwent hyperparameter optimization using first RandomizedSearchCV and then a bounded optimization using the GridSearchCV method,

both from the sklearn library. The hyperparameters and ranges of variation are presented in Table 4-1.

Table 4-1. Hyperparameters used to train the models

Models	Hyperparameters	Values
Random Forest	criterion	['squared_error', 'absolute_error', 'friedman_mse']
	n_estimators	range (10,500,50)
	max_depth	None + range (2,30,5)
	min_samples_split	Range (2,30,5)
	min_samples_leaf	[1, 2, 4]
	max_features	[1.0, 'sqrt', 'log2']
Extra Trees	criterion	['squared_error', 'absolute_error', 'friedman_mse']
	n_estimators	range (10,500,50)
	max_depth	None + range (2,30,5)
	min_samples_split	Range (2,30,5)
	min_samples_leaf	[1, 2, 4]
	max_features	[1.0, 'sqrt', 'log2']
Gradient Boost Regressor Tree	criterion	['squared_error', 'friedman_mse']
	n_estimators	range (10,500,50)
	learning_rate	[0.01, 0.1, 0.2, 0.5]
	max_depth	[2-10]
	subsample	[0.7, 0.8, 0.9]
Extreme Gradient Boost	n_estimators	range (10,500,50)
	learning_rate	[0.01, 0.1, 0.2, 0.5]
	max_depth	[2-10]
	subsample	[0.7, 0.8, 0.9]
	colsample_bytree	[0.7, 0.8, 0.9]
	gamma	[0, 0.1, 0.2]
	reg_alpha	[0, 0.001, 0.005]
	reg_lambda	[1, 1.5, 2]

To evaluate and compare the performance of the implemented regression models, three key metrics were selected based on the systematic literature review: the mean absolute error (MAE), the root mean square error (RMSE), and the coefficient of determination (R^2). Both MAE and RMSE quantify the average prediction error, but they differ in their sensitivity to outliers. While MAE represents the straightforward average distance between predicted and actual values, RMSE penalizes larger errors more significantly due to its squaring mechanism. [80] In parallel, the R^2 value assesses the quality of the model by indicating the proportion of variance that it can explain, offering a more intuitive measure of its predictive

accuracy. [81] To complement these quantitative measures, deviation curves, residual plots, and learning curves were constructed. These visualizations allow for a direct assessment of the model's performance and the distribution of errors. The combined interpretation of these statistical metrics and visual plots provides a wider analysis of each model's predictive capability for the given dataset.

5. Results

This chapter presents the prediction of interfacial tension in nanofluids containing surfactants. It begins by detailing the experimentally determined IFT values from an oil-water system formulated with a surfactant and four distinct carbon-based nanomaterials. Subsequently, the chapter provides an exploratory analysis of the dataset, examining data distribution and correlations to establish the relationships between input variables and the resulting IFT. The final section evaluates the performance of the machine learning models, including an analysis of key performance metrics, a comparison of predicted versus actual values, a review of the model residuals, and learning curves.

5.1 Interfacial Tension Measurements of Nanofluids with Surfactants

Figure 5-1 compares the interfacial tension values for the four different carbon quantum dots (CQDs) as a function of concentration, presenting results in the absence of a surfactant. As shown in the figure, the addition of CQDs generally reduces the IFT from its initial value. This reduction is attributed to two key properties of the nanoparticles: their surface contains functional groups that interact with water molecules to weaken cohesive forces, [10] and their high surface-area-to-volume ratio provides numerous sites for molecular interaction at

the interface. [82] However, this effect is highly dependent on concentration. At high concentrations, particularly for the commercial blue and molasses derived CQDs, the trend reverses and IFT increases, likely due to surface saturation that promotes particle aggregation. Despite this limitation, it is noteworthy that the CQDs synthesized from molasses and coffee were significantly more efficient at lowering IFT than their commercial counterparts, with the molasses-derived CQDs demonstrating the best overall performance.

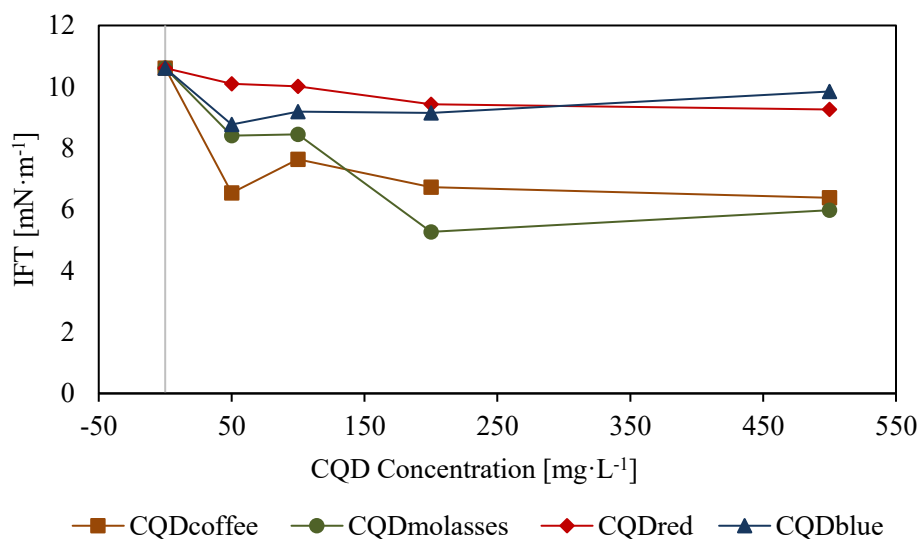


Figure 5-1. IFT vs Concentration CQD

Figure 5.2 illustrates the changes in interfacial tension as a function of the concentration of the four nanomaterials, commercial blue and red, molasses-derived, and coffee-derived, all in the presence of a surfactant. The results reveal a variable pattern, indicating a complex interaction between the CQDs and the surfactant molecules at the interface. Notably, the commercial CQDs caused an increase in IFT, in contrast to the CQDs synthesized from molasses and coffee, which continued to reduce it. Among these, the molasses-derived CQDs delivered the best performance by achieving the most significant IFT reduction. However, at higher concentrations, their effect began to resemble to the coffee-derived CQDs. These variations can be attributed to modifications in the interfacial morphology and adsorption dynamics. Such changes can lead to faster and more stable interfacial coverage, with the CQDs providing additional mechanical stabilization to the surfactant film. [83]

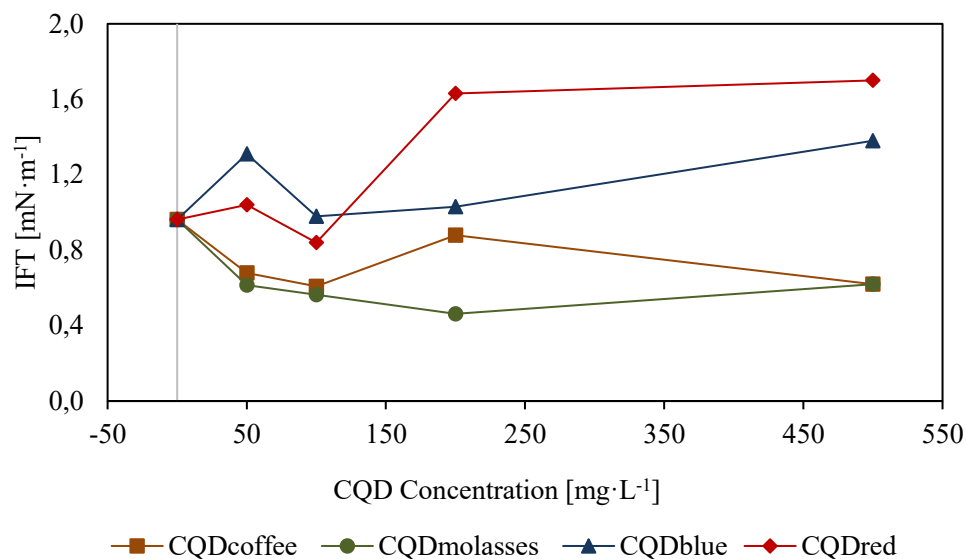


Figure 5-2. IFT vs Concentration CQD in the presence of Surfactant

5.2 Analysis and Description of the Dataset's Structure

A representative dataset comprising 551 samples and 11 variables was constructed by combining original experimental measurements with data from bibliographic sources. This dataset then underwent a descriptive analysis to determine its distribution. As illustrated in the distribution plot in *Figure 5-3*, the data are heavily skewed toward lower values. The distribution is highly concentrated below $20.55 \text{ mN}\cdot\text{m}^{-1}$, with 50% of the data falling below $7.50 \text{ mN}\cdot\text{m}^{-1}$. While this indicates a significant imbalance in the dataset, as confirmed with the information in *Table 5-1*, the values in the upper quartile are not outliers. Instead, these higher values represent the crucial baseline interfacial tension of the initial system before the addition of surfactants and nanomaterials.

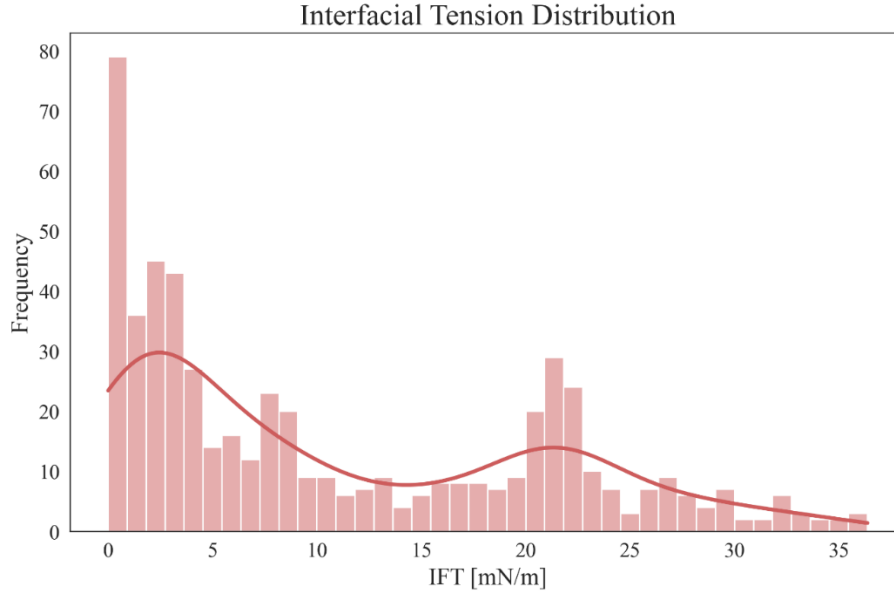


Figure 5-3. Dataset distribution

Table 5-1. Descriptive Statistics of Interfacial Tension in the Dataset

IFT	Mean	Std	Min	25%	50%	75%	Max
	10.76	9.85	0.0001	2.28	7.50	20.55	36.36

Additionally, to understand the relationships between the system variables, a Spearman correlation analysis was conducted. The resulting correlation matrix is visualized as a heatmap in *Figure 5-4*, where red and blue shades indicate positive and negative correlations, respectively. The analysis reveals several key relationships. A very strong positive correlation ($\rho = 0.9$) was found between HLB and surfactant concentration, which is consistent with the phenomenon. Most importantly for this study, a strong negative correlation was observed between interfacial tension and surfactant concentration ($\rho = -0.7$), as expected, highlighting the latter's primary role in reducing IFT. Furthermore, HLB ($\rho = -0.6$), API^o ($\rho = -0.5$), and nanomaterial concentration ($\rho = -0.4$) also demonstrated significant negative correlations with IFT, indicating that although they are not the most important factor, they have a significant effect on reducing interfacial tension.

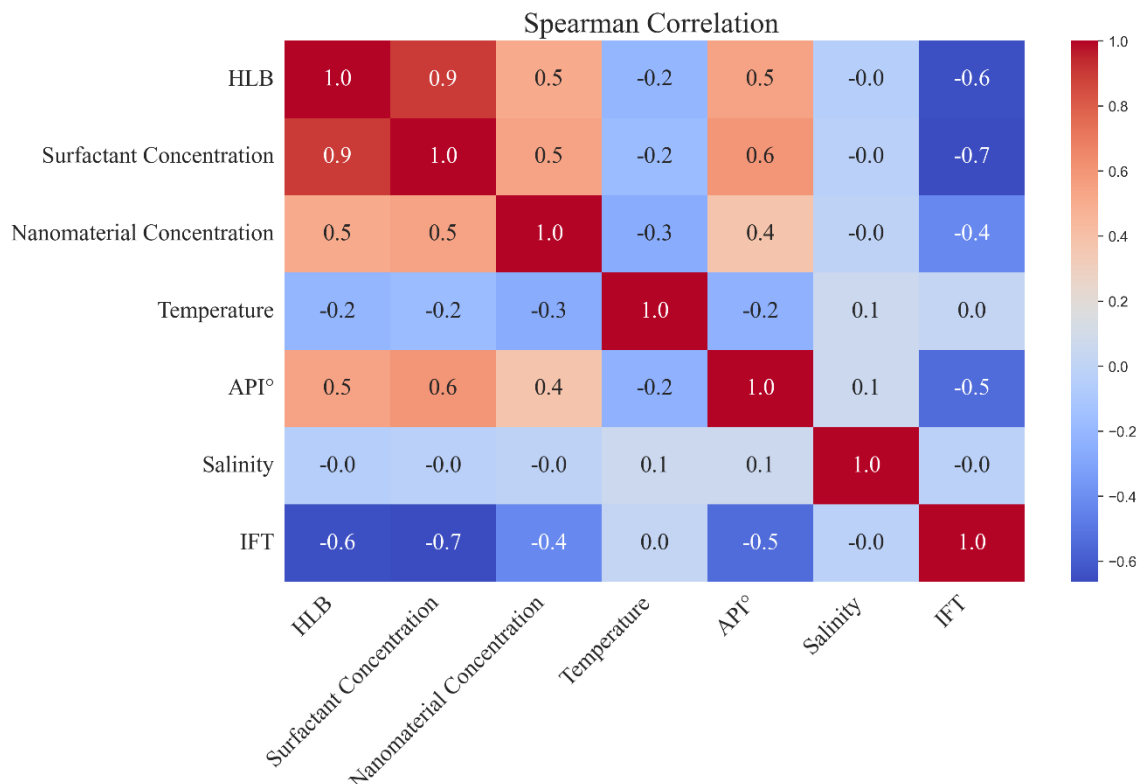


Figure 5-4. Spearman correlation matrix for the continuous variables.

5.3 Performance of the Implemented Machine Learning Models

An evaluation of the models for predicting interfacial tension revealed the disparities in performance. Based on the validation metrics illustrated in *Figure 5-5*, the random forest (RF) model demonstrated superior accuracy. It achieved the highest coefficient of determination ($R^2 = 0.854$) and the lowest root mean square error (RMSE = 3.46). The gradient boosting (GBRT) model also performed well, yielding the lowest mean absolute error (MAE = 1.85) and the second-highest R^2 score (0.832), positioning it as a strong alternative. In contrast, the XGBoost and extra trees (ET) models were less effective. The extra trees algorithm yielded the weakest predictions, exhibiting the highest error rates (RMSE= 4.01, MAE = 2.08) and the lowest R^2 value (0.803). These results suggest that for this specific task, the random forest architecture most effectively captured the data's underlying patterns and variable correlations.

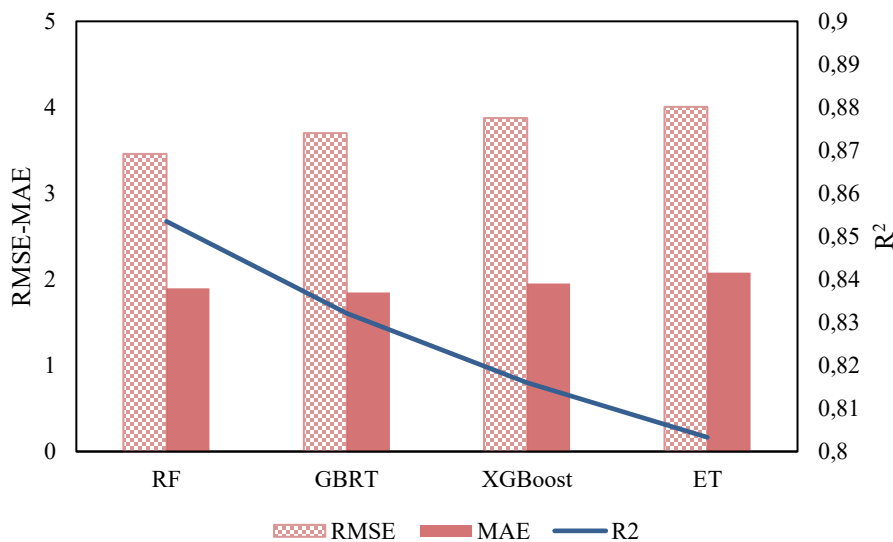


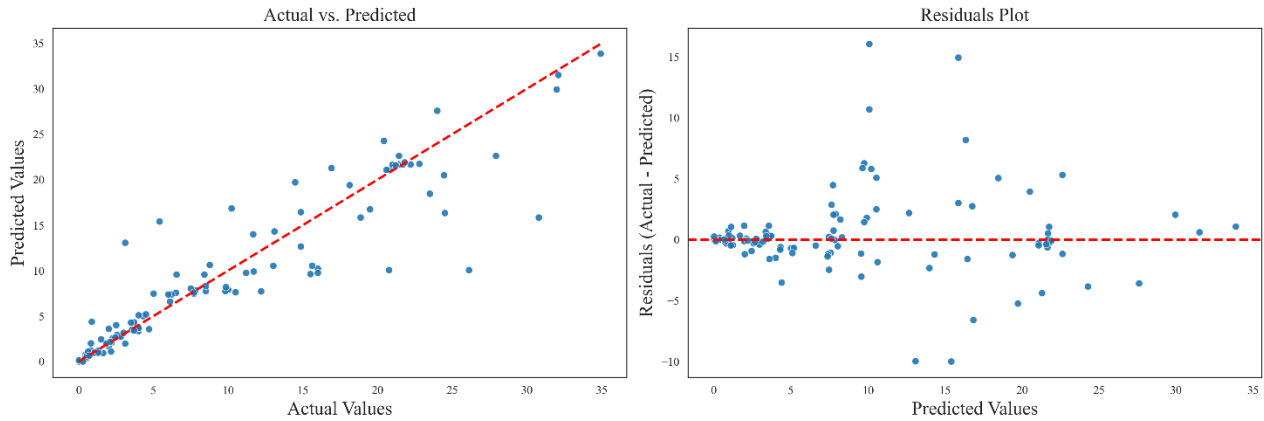
Figure 5-5. Performance metrics.

Figure 5-6 provides a visual assessment of the final tuned random forest (RF) and extra trees (ET) models. The "Actual vs. Predicted" scatter plots (left panels) illustrate the correlation between the true interfacial tension values and the model's predictions. In both cases, the data points generally cluster around the diagonal line of perfect prediction ($y=x$), indicating a strong positive correlation, especially for interfacial tension values below $15 \text{ mN}\cdot\text{m}^{-1}$. However, a notable limitation is evident in both models. They exhibit a performance ceiling, consistently struggling to predict values above approximately $25 \text{ mN}\cdot\text{m}^{-1}$, often underestimating samples where the true value is higher. This issue is likely attributable to an imbalance in the dataset, where a scarcity of high-value training instances prevented the models from effectively learning the patterns associated with high interfacial tension.

Further insights are revealed in the residual plots (right panels), which display the prediction error (Actual - Predicted) against the predicted value. An ideal model would show residuals randomly scattered around the zero-error line. Instead, both the RF and ET plots exhibit clear heteroscedasticity, a funnel shape where the variance of the error increases as the predicted value gets larger. This pattern confirms that both models are significantly less reliable and produce larger errors when predicting higher values of interfacial tension. While this behavior is present in both, the extra trees model shows slightly greater error dispersion and several larger individual residuals, consistent with its quantitative metrics.

A

Analysis of RandomForest Model



B

Analysis of ExtraTrees Model

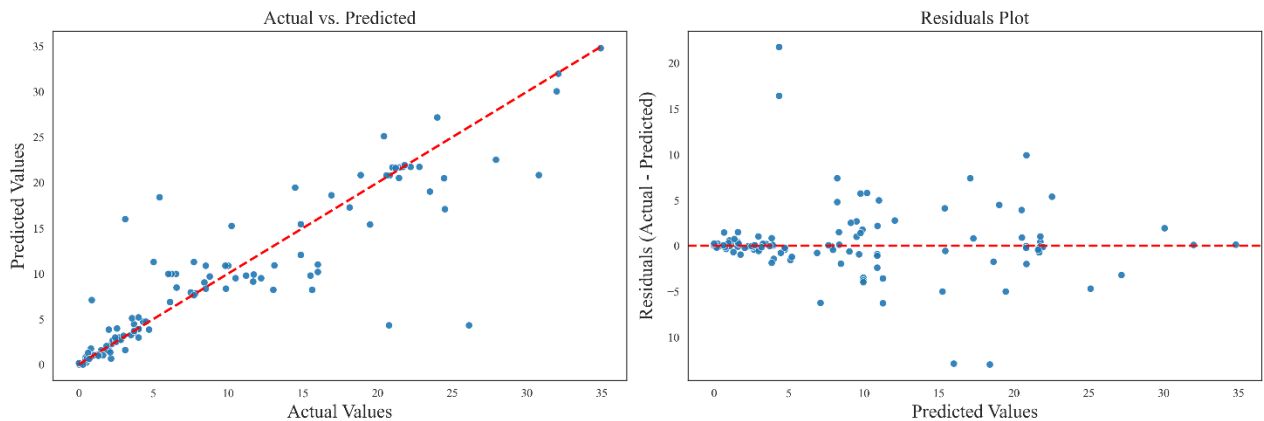


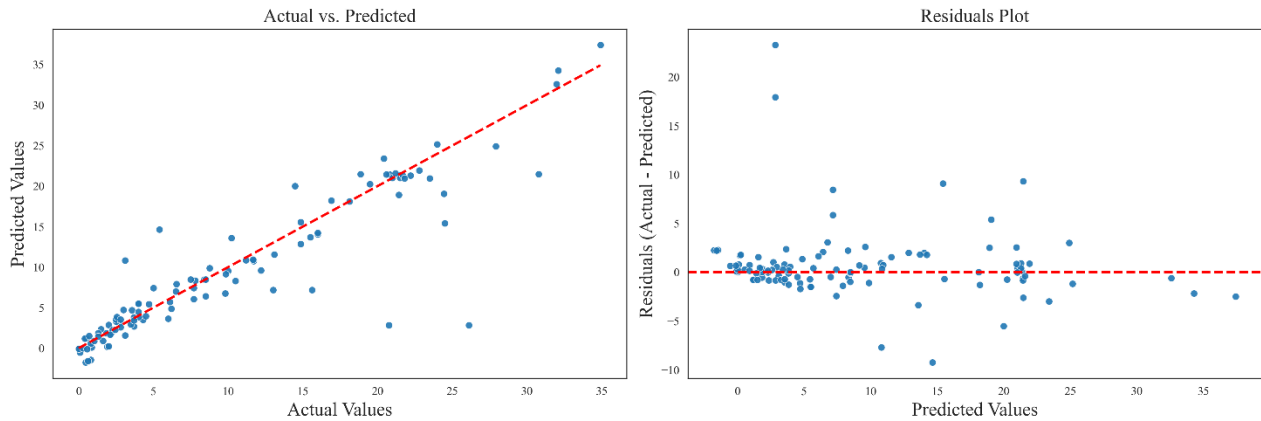
Figure 5-6. Evaluation of Bagging Models: Actual vs. Predicted Values and Residual Plots for (A) Random Forest and (B) Extra Trees

Similarly, the boosting models, gradient boosting (GBRT) and XGBoost, exhibit the same performance challenges identified in the bagging models. Their prediction plots, presented in Figure 5-7, also show a decline in accuracy for high values, and their residual plots confirm the presence of heteroscedasticity. However, a closer inspection of their residuals reveals a more systematic bias. Unlike the bagging models, both GBRT and XGBoost show a strong tendency to underpredict high-value samples, as evidenced by the prevalence of large positive residuals. This indicates that while they share the same general limitations,

their error patterns are less symmetric, pointing to a specific weakness in estimating the upper range of the data spectrum.

A

Analysis of GBRT Model



B

Analysis of XGBoost Model

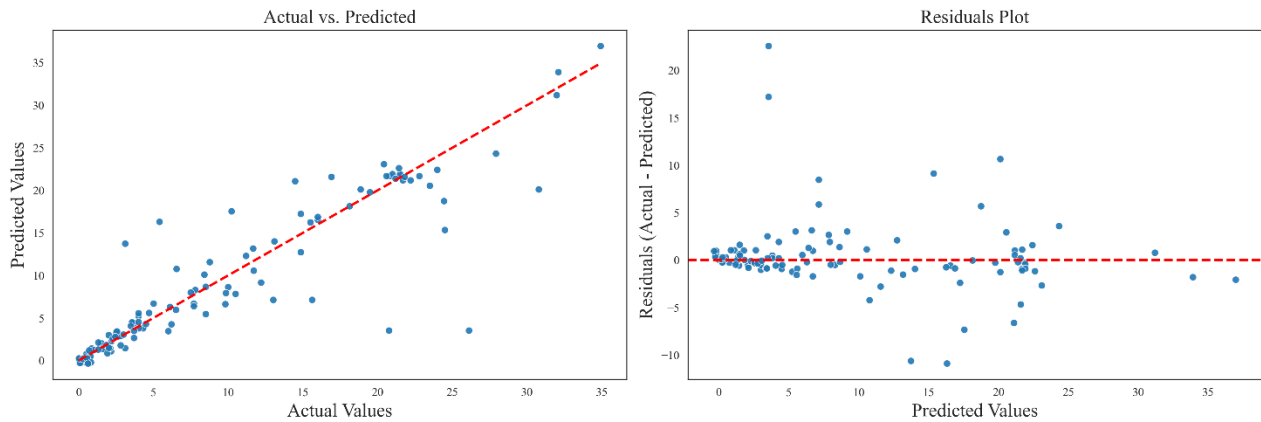


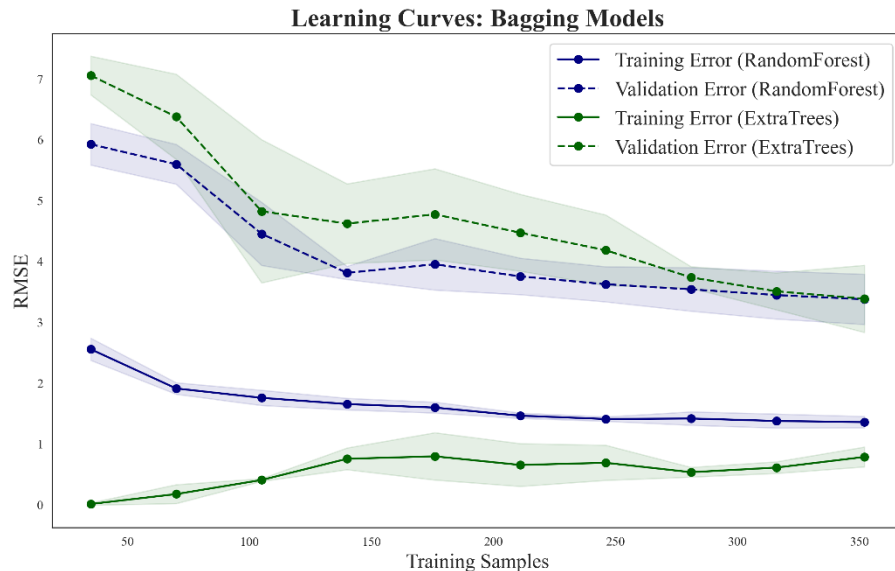
Figure 5-7. Evaluation of Boosting Models: Actual vs. Predicted Values and Residual Plots for (A) Gradient Boost Regressor Tree and (B) Extreme Gradient Boost

To further diagnose model behavior, *Figure 5-8* presents the learning curves, which plot the model's performance (RMSE) against the number of training samples. These curves are essential for evaluating generalization and identifying issues such as overfitting (high variance) or underfitting (high bias). A distinct difference in learning behavior is observed between the two bagging models. The random forest (RF) model demonstrates excellent generalization. Its training and validation error curves converge steadily, resulting in a small

final gap. This indicates that the model learns meaningful patterns from the data without significant overfitting, which explains its superior performance in the quantitative metrics. In contrast, the extra trees model exhibits signs of high variance. Its training error remains near-zero while a large gap persists between it and the validation error, showing that it memorized the training data rather than learning generalizable patterns.

The two boosting models (GBRT and XGBoost) display nearly identical learning curves. Similar to extra trees, they achieve very low training error, suggesting a tendency to overfit. While their validation errors decrease sharply and ultimately settle at a lower RMSE than the bagging models, a significant generalization gap remains. The curves also show that their performance stops improving after approximately 200 training samples, indicating that adding more data becomes inefficient. These learning curves confirm that the random forest model is the most balanced and reliable. It is the only model that effectively avoids significant overfitting, leading to the best generalization performance on unseen data.

A



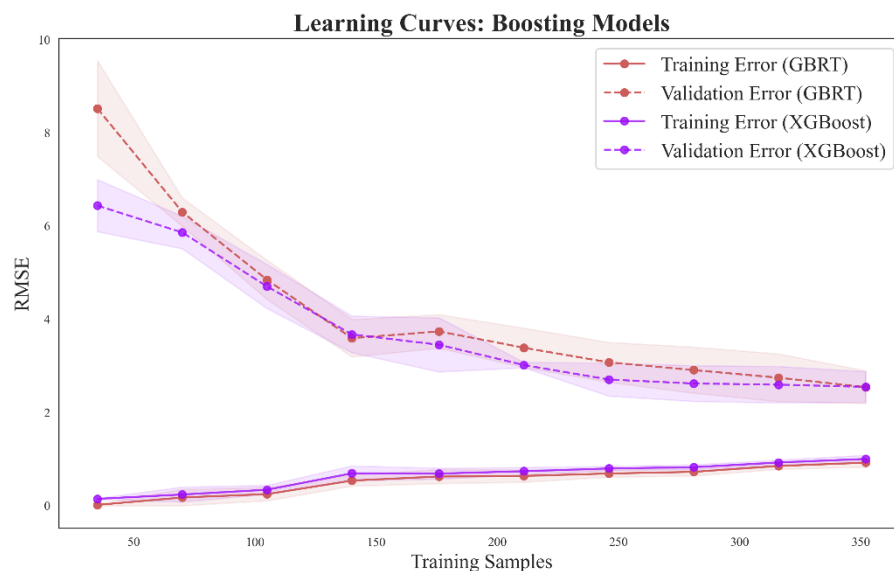
B

Figure 5-8. Learning Curves for Bagging and Boosting Models using RMSE: (A) Extra Trees and Random Forest, and (B) Gradient Boosting and XGBoost.

The optimized hyperparameters presented in *Table 5-2* provide insight into how each model adapted to the specific challenges of the interfacial tension (IFT) dataset, particularly its imbalanced distribution with a scarcity of high-value samples. For random forest and extra trees, the selection of a moderately deep tree structure (`max_depth: 12`) indicates that accurately predicting IFT requires capturing complex, high-order interactions between the input features. The models needed this level of complexity to differentiate the subtle conditions that lead to a wide range of IFT values. This depth represents a critical balance found during tuning deep enough to be expressive, yet constrained enough to prevent memorizing the few high-value data points.

In the same way, the boosting models adopted a strategy of combining many simple rules, evidenced by their use of very shallow trees (`max_depth: 2`). This suggests an iterative approach where the final IFT prediction is built by aggregating the effects of numerous basic conditions present in the data. The high learning rate (~ 0.5) enabled fast error correction. At the same time, crucial parameters such as `subsample: 0.7` and XGBoost's explicit regularization were essential for preventing the models from focusing solely on the dominant, low-IFT data region, thereby improving generalization.

Table 5-2. Tuned Model Hyperparameters

Models	Hyperparameters	Values
Random Forest	criterion	'squared_error'
	n_estimators	320
	max_depth	12
	min_samples_split	2
	min_samples_leaf	1
	max_features	1
Extra Trees	criterion	'squared_error'
	n_estimators	150
	max_depth	12
	min_samples_split	2
	min_samples_leaf	1
	max_features	1
Gradient Boost Regressor Tree	criterion	'squared_error'
	n_estimators	410
	learning_rate	0.5
	max_depth	2
	subsample	0.7
Extreme Gradient Boost	n_estimators	450
	learning_rate	0.45
	max_depth	2
	subsample	0.7
	colsample_bytree	0.9
	gamma	0.05
	reg_alpha	0.006
	reg_lambda	1

Ultimately, these configurations demonstrate that for this dataset, the random forest's strategy of averaging predictions from complex, deep-tree models proved more effective at navigating data imbalance, yielding a more robust predictor than the sequential, error-correcting approach of the boosting models.

6. Conclusions and Future Work

This work demonstrated the successful implementation of machine learning models for predicting interfacial tension (IFT) in complex chemical systems designed for enhanced oil recovery (eOR). An experimental evaluation was conducted on a crude oil-brine system from a Colombian field, assessing the effect of surfactants combined with four distinct carbon quantum dots (CQDs). To enrich the dataset, these experimental results were then supplemented with relevant data collected from a review of existing literature. The resulting data were used to create a representative dataset to train and compare four supervised regression models, with key variables including the nanomaterial properties (type and concentration), surfactant properties (type, concentration, and hydrophilic-lipophilic balance), and system conditions (brine salinity, crude oil °API, and temperature).

The analysis established that all models faced a significant challenge due to the dataset's imbalanced distribution, with a majority of IFT values concentrated below $20.55 \text{ mN}\cdot\text{m}^{-1}$. This scarcity of high-value data points led to heteroscedasticity in the model predictions, where error margins increased for higher IFT values. The machine learning models presented varied performance in navigating this challenge. While all models captured the general trends, the random forest (RF) model consistently outperformed the others, achieving the highest R^2 and the lowest RMSE and MAE metrics.

The superior performance of the random forest model is attributed to its excellent generalization capabilities, as confirmed by the learning curve analysis. Unlike the extra trees and boosting models, which exhibited clear signs of overfitting, the RF model's training and validation error curves converged effectively, indicating a well-balanced model. This result suggests that the RF's strategy of averaging predictions from multiple, deep-tree ensembles was more robust for this imbalanced dataset than the sequential, error-correcting approach of the boosting models. These findings position ensemble methods like random forest as powerful tools for applications where the dataset is small and imbalanced, yet contains outliers that are critical to the physical phenomenon being studied.

Although the learning models produced predictions close to the actual values, it is clear that the dataset must be expanded to improve the error metrics and address the data imbalance. This study serves as a foundational framework, demonstrating that ML can be a powerful complementary tool for extrapolating experimental results and accelerating the screening process for cEOR formulations.

The results obtained in this research can also be used to establish several lines of research for future work:

- Expansion and diversification of the dataset: To improve the accuracy and generalization capabilities of the predictive models, it is essential to expand the training base. Future work should focus on incorporating more data, particularly in the high-IFT range.
- Evaluation in dynamic and porous media: While this study focused on static IFT measurements, future research should extend to dynamic testing. Evaluating the performance of the top-performing chemical formulations in core-flooding experiments would be essential to validate their effectiveness under conditions representative of a porous medium and to establish a direct link between IFT reduction and improvements in the amount of oil recovered.

7. References

- [1] T. Castillo, Garcia, F., Mosquera, L., Rivadeneira, T., Segura, K., Yujato, M., "Panorama Energético de America Latina y El Caribe," Organización Latinoamericana de Energía, olade.org2021.
- [2] IEA. (2023). *Where does Colombia get its energy?* Available: <https://www.iea.org/countries/colombia/energy-mix#where-does-colombia-get-its-energy>
- [3] "Informe Hechos de Sostenibilidad," Asociación Colombiana del Petróleo y Gas2024.
- [4] F. Umeadi and Y. TPuyate, "Review of Enhanced Oil Recovery Methods and the way Forward," *International Journal of Advances in Engineering and Management (IJAEM)*, vol. 4, p. 8, 2022.
- [5] L. J. Roldán Vargas, "Desarrollo de nanomateriales para la liberación controlada de surfactantes en recobro mejorado de petróleo."
- [6] M. Vinod, P. Jain, and E. Prasad. (2023). *Chemical Enhanced Oil Recovery (EOR) Market by Type (Surfactant, Polymer, and ASP), Application (Onshore and Offshore): Global Opportunity Analysis and Industry Forecast, 2022–2031*. Available: <https://www.alliedmarketresearch.com/chemical-enhanced-oil-recovery-EOR-market>
- [7] J. J. Sheng, *Modern chemical enhanced oil recovery: theory and practice*: Gulf Professional Publishing, 2010.
- [8] J. J. Sheng, "A comprehensive review of alkaline-surfactant-polymer (ASP) flooding," in *SPE Western Regional Meeting*, 2013, pp. SPE-165358-MS.
- [9] C. A. Franco, C. H. Candela, J. Gallego, J. Marin, L. E. Patiño, N. Ospina, *et al.*, "Easy and rapid synthesis of carbon quantum dots from Mortino (*Vaccinium Meridionale Swartz*) extract for use as green tracers in the oil and gas industry: Lab-to-field trial development in Colombia," *Industrial & Engineering Chemistry Research*, vol. 59, pp. 11359-11369, 2020.
- [10] M. AfzaliTabar, A. Rashidi, M. Alaei, H. Koolivand, S. Pourhashem, and S. Askari, "Hybrid of quantum dots for interfacial tension reduction and reservoir alteration wettability for enhanced oil recovery (EOR)," *Journal of molecular liquids*, vol. 307, p. 112984, 2020.
- [11] F. Salimi, E. Jafarbeigi, C. Karami, and E. Khodapanah, "Synthesis of cost-effective Si-CQD for effective oil separation from core rock," *Journal of Molecular Liquids*, vol. 394, p. 123722, 2024.
- [12] I.-A. Baragau, Z. Lu, N. P. Power, D. J. Morgan, J. Bowen, P. Diaz, *et al.*, "Continuous hydrothermal flow synthesis of S-functionalised carbon quantum dots for enhanced oil recovery," *Chemical Engineering Journal*, vol. 405, p. 126631, 2021.
- [13] M. Mirzavandi, J. A. Ali, A. K. Manshad, B. Majeed, B. S. Mahmood, A. H. Mohammadi, *et al.*, "Performance evaluation of silica-graphene quantum dots for enhanced oil recovery from carbonate reservoirs," *Energy & Fuels*, vol. 37, pp. 955-964, 2023.
- [14] M. R. Aghajanzadeh, P. Ahmadi, M. Sharifi, and M. Riazi, "Wettability modification of oil-wet carbonate reservoirs using silica-based nanofluid: An experimental approach," *Journal of Petroleum Science and Engineering*, vol. 178, pp. 700-710, 2019.
- [15] H. Rezvani, A. Khalilnejad, and A. A. Sadeghi-Bagherabadi, "Comparative experimental study of various metal oxide nanoparticles for the wettability alteration of carbonate rocks in EOR processes," in *80th EAGE Conference and Exhibition 2018*, 2018, pp. 1-5.
- [16] G. Dordzie and M. Dejam, "Enhanced oil recovery from fractured carbonate reservoirs using nanoparticles with low salinity water and surfactant: A review on experimental and simulation studies," *Advances in Colloid and Interface Science*, vol. 293, p. 102449, 2021.
- [17] S. O. Olayiwola and M. Dejam, "A comprehensive review on interaction of nanoparticles with low salinity water and surfactant for enhanced oil recovery in sandstone and carbonate reservoirs," *Fuel*, vol. 241, pp. 1045-1057, 2019.

- [18] Y. Cheraghi, S. Kord, and V. Mashayekhizadeh, "Application of machine learning techniques for selecting the most suitable enhanced oil recovery method; challenges and opportunities," *Journal of Petroleum Science and Engineering*, vol. 205, p. 108761, 2021.
- [19] R. He, M. Weizhong, X. Ma, and Y. Liu, "Modeling and optimizing for operation of CO₂-EOR project based on machine learning methods and greedy algorithm," *Energy Reports*, vol. 7, pp. 3664-3677, 2021.
- [20] S. Céspedes, A. Molina, B. Lerner, M. S. Pérez, C. A. Franco, and F. B. Cortés, "A selection flowchart for micromodel experiments based on computational fluid dynamic simulations of surfactant flooding in enhanced oil recovery," *Processes*, vol. 9, p. 1887, 2021.
- [21] B. A. Suleimanov, F. S. Ismayilov, O. A. Dyshin, and E. F. Veliyev, "Selection methodology for screening evaluation of EOR methods," *Petroleum Science and Technology*, vol. 34, pp. 961-970, 2016.
- [22] S. Su, N. Zhang, P. Wang, S. Jia, A. Zhang, H. Wang, *et al.*, "Investigation and optimization of EOR screening by implementing machine learning algorithms," *Applied Sciences*, vol. 13, p. 12267, 2023.
- [23] ANH. (2024). *Simposio EOR: Un camino para maximizar reservas*. Available: <https://www.anh.gov.co/es/noticias/simposio-eor-un-camino-para-maximizar-reservas/>
- [24] M. S. Azad, "IFT role on oil recovery during surfactant based EOR methods," in *Surfactants in Upstream E&P*, ed: Springer, 2021, pp. 115-148.
- [25] X.-J. Zhang, Z.-H. Zhou, L. Han, Y.-Q. Zhang, Q. Zhang, D.-S. Ma, *et al.*, "Mechanism responsible for the reduction of interfacial tension by extended surfactants," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 634, p. 128013, 2022.
- [26] N. S. Muhammed, T. Olayiwola, S. Elkatatny, B. Haq, and S. Patil, "Insights into the application of surfactants and nanomaterials as shale inhibitors for water-based drilling fluid: A review," *Journal of Natural Gas Science and Engineering*, vol. 92, p. 103987, 2021.
- [27] S. R. Oscar E. Medina, Nathaly Garzón, Daniel López, Esteban A. Taborda, Juan C. Ordoñez, Sergio Augusto Fernández, Farid B. Cortés, Camilo A. Franco, "Advances in Quantum Dot Applications for the Oil and Gas Industry: Current Trends and Future Directions," *ACS*, p. 39, 2024.
- [28] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, pp. 381-386, 2020.
- [29] A. K. Idogun, R. O. Ujah, and L. A. James, "Surrogate-Based Analysis of Chemical Enhanced Oil Recovery—A Comparative Analysis of Machine Learning Model Performance," in *SPE Nigeria Annual International Conference and Exhibition*, 2021, p. D021S006R006.
- [30] A. Bonnet. (2023, 04/07/2025). *What is Ensemble Learning?* Available: <https://encord.com/blog/what-is-ensemble-learning/>
- [31] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, pp. 197-227, 2016.
- [32] A. Parmar, R. Katariya, and V. Patel, "A review on random forest: An ensemble classifier," in *International conference on intelligent data communication technologies and internet of things*, 2018, pp. 758-763.
- [33] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, pp. 3-42, 2006.
- [34] C. Bentéjac, A. Csörgö, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, pp. 1937-1967, 2021.
- [35] M. F. El-Amin, B. Alwated, and H. A. Hoteit, "Machine learning prediction of nanoparticle transport with two-phase flow in porous media," *Energies*, vol. 16, p. 678, 2023.

- [36] B. Wei, Y. He, J. You, S. Wen, and J. Tang, "Interpretable Machine Learning for Prediction of Minimum Miscibility Pressure in CO₂-Oil System Considering Nano-Confinement Effect," in *International Petroleum Technology Conference*, 2024, p. D021S053R005.
- [37] H. Guo, M. Dou, W. Hanqing, F. Wang, G. Yuanyuan, Z. Yu, *et al.*, "Proper use of capillary number in chemical flooding," *Journal of Chemistry*, vol. 2017, p. 4307368, 2017.
- [38] Y. A. Alzahid, P. Mostaghimi, S. D. Walsh, and R. T. Armstrong, "Flow regimes during surfactant flooding: The influence of phase behaviour," *Fuel*, vol. 236, pp. 851-860, 2019.
- [39] X. Hou and J. J. Sheng, "Experimental study on the imbibition mechanism of the Winsor type I surfactant system with ultra-low IFT in oil-wet shale oil reservoirs by NMR," *Journal of Petroleum Science and Engineering*, vol. 216, p. 110785, 2022.
- [40] N. Kumar and A. Mandal, "Surfactant stabilized oil-in-water nanoemulsion: stability, interfacial tension, and rheology study for enhanced oil recovery application," *Energy & fuels*, vol. 32, pp. 6452-6466, 2018.
- [41] B. Huang, X. Li, W. Zhang, C. Fu, Y. Wang, and S. Fu, "Study on demulsification-flocculation mechanism of oil-water emulsion in produced water from alkali/surfactant/polymer flooding," *Polymers*, vol. 11, p. 395, 2019.
- [42] V. Alvarado, X. Wang, and M. Moradi, "Stability proxies for water-in-oil emulsions and implications in aqueous-based enhanced oil recovery," *Energies*, vol. 4, pp. 1058-1086, 2011.
- [43] S. Solairaj, C. Britton, D. H. Kim, U. Weerasooriya, and G. A. Pope, "Measurement and analysis of surfactant retention," in *SPE Improved Oil Recovery Conference?*, 2012, pp. SPE-154247-MS.
- [44] C. A. Franco, L. J. Giraldo, C. H. Candela, K. M. Bernal, F. Villamil, D. Montes, *et al.*, "Design and tuning of nanofluids applied to chemical enhanced oil recovery based on the surfactant-nanoparticle-brine interaction: From laboratory experiments to oil field application," *Nanomaterials*, vol. 10, p. 1579, 2020.
- [45] N. K. Maurya and A. Mandal, "Investigation of synergistic effect of nanoparticle and surfactant in macro emulsion based EOR application in oil reservoirs," *Chemical Engineering Research and Design*, vol. 132, pp. 370-384, 2018.
- [46] N. Yekeen, E. Padmanabhan, A. H. Syed, T. Sevoo, and K. Kanesen, "Synergistic influence of nanoparticles and surfactants on interfacial tension reduction, wettability alteration and stabilization of oil-in-water emulsion," *Journal of Petroleum Science and Engineering*, vol. 186, p. 106779, 2020.
- [47] Z. Liu, V. Bode, P. Hadayati, H. Onay, and E. J. Sudhölter, "Understanding the stability mechanism of silica nanoparticles: The effect of cations and EOR chemicals," *Fuel*, vol. 280, p. 118650, 2020.
- [48] E. Nourafkan, Z. Hu, M. Garum, H. Esmaeili, and D. Wen, "Nanomaterials for subsurface application: study of particles retention in porous media," *Applied Nanoscience*, vol. 11, pp. 1847-1856, 2021.
- [49] E. Nourafkan, Z. Hu, and D. Wen, "Nanoparticle-enabled delivery of surfactants in porous media," *Journal of colloid and interface science*, vol. 519, pp. 44-57, 2018.
- [50] N. Akhlaghi, S. Riahi, and R. Parvaneh, "Interfacial tension behavior of a nonionic surfactant in oil/water system; salinity, pH, temperature, and ionic strength effects," *Journal of Petroleum Science and Engineering*, vol. 198, p. 108177, 2021.
- [51] Y. Wei and T. Babadagli, "Alteration of interfacial properties by chemicals and nanomaterials to improve heavy oil recovery at elevated temperatures," *Energy & Fuels*, vol. 31, pp. 11866-11883, 2017.
- [52] H. Divandari, A. Hemmati-Sarapardeh, M. Schaffie, and M. Ranjbar, "Integrating functionalized magnetite nanoparticles with low salinity water and surfactant solution: Interfacial tension study," *Fuel*, vol. 281, p. 118641, 2020.

-
- [53] B. N. Nwani, M. S. Azad, and J. Trivedi, "Effect of various classes of surfactants on interfacial tension reduction and wettability alteration on smart-water-surfactant systems," *Energy & Fuels*, vol. 36, pp. 251-261, 2021.
- [54] M. Almahfood and B. Bai, "The synergistic effects of nanoparticle-surfactant nanofluids in EOR applications," *Journal of Petroleum Science and Engineering*, vol. 171, pp. 196-210, 2018.
- [55] M. S. Kamal, I. A. Hussein, and A. S. Sultan, "Review on surfactant flooding: phase behavior, retention, IFT, and field applications," *Energy & fuels*, vol. 31, pp. 7701-7720, 2017.
- [56] M. N. Bello and A. Shafiei, "A novel green nanocomposite for EOR: Experimental Investigation of IFT Reduction, wettability Shift, and nanofluid stability," *Journal of Molecular Liquids*, vol. 414, p. 126187, 2024.
- [57] W. S. Bainbridge and M. C. Roco, "Science and technology convergence: with emphasis for nanotechnology-inspired convergence," *Journal of Nanoparticle Research*, vol. 18, p. 211, 2016.
- [58] A. Bigdeli and M. Delshad, "The evolving landscape of oil and gas chemicals: convergence of artificial intelligence and chemical-enhanced oil recovery in the energy transition toward sustainable energy systems and net-zero emissions," *Journal of Data Science and Intelligent Systems*, vol. 2, pp. 65-78, 2024.
- [59] M. Vrabel, "Preferred reporting items for systematic reviews and meta-analyses," *Number 5/September 2015*, vol. 42, pp. 552-554, 2015.
- [60] A. A. Ivanova, C. Phan, A. Barifciani, S. Iglauer, and A. N. Cheremisin, "Effect of nanoparticles on viscosity and interfacial tension of aqueous surfactant solutions at high salinity and high temperature," *Journal of Surfactants and Detergents*, vol. 23, pp. 327-338, 2020.
- [61] A. Rashidi-Khaniabadi, E. Rashidi-Khaniabadi, B. Amiri-Ramsheh, M.-R. Mohammadi, and A. Hemmati-Sarapardeh, "Modeling interfacial tension of surfactant–hydrocarbon systems using robust tree-based machine learning algorithms," *Scientific Reports*, vol. 13, p. 10836, 2023.
- [62] A. Gazem, H. Patel, H. Sreenivasan, C. Sahu, and S. Krishna, "Combined effect of silica nanoparticles and binary surfactants in enhancing oil recovery: an experimental investigation," *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, vol. 702, p. 134980, 2024.
- [63] S. Betancur, L. J. Giraldo, F. Carrasco-Marín, M. Riazi, E. J. Manrique, H. Quintero, *et al.*, "Importance of the nanofluid preparation for ultra-low interfacial tension in enhanced oil recovery based on surfactant–nanoparticle–brine system interaction," *ACS omega*, vol. 4, pp. 16171-16180, 2019.
- [64] M. Zallaghi, R. Kharrat, and A. Hashemi, "Improving the microscopic sweep efficiency of water flooding using silica nanoparticles," *Journal of Petroleum Exploration and Production Technology*, vol. 8, pp. 259-269, 2018.
- [65] M. Mohajeri, M. R. Rasaei, and M. Hekmatzadeh, "Experimental study on using SiO₂ nanoparticles along with surfactant in an EOR process in micromodel," *Petroleum Research*, vol. 4, pp. 59-70, 2019.
- [66] K. R. Aurand, G. S. Dahle, and O. Torsæter, "Comparison of oil recovery for six nanofluids in Berea sandstone cores," in *the International Symposium of the Society of Core Analysts held in Avignon, France*, 2014.
- [67] M. Mohajeri, M. Hemmati, and A. S. Shekarabi, "An experimental study on using a nanosurfactant in an EOR process of heavy oil in a fractured micromodel," *Journal of petroleum Science and engineering*, vol. 126, pp. 162-173, 2015.

- [68] F. Yousefmarzi, A. Haratian, J. Mahdavi Kalatehno, and M. Keihani Kamal, "Machine learning approaches for estimating interfacial tension between oil/gas and oil/water systems: a performance analysis," *Scientific Reports*, vol. 14, p. 858, 2024.
- [69] A. Talapatra, B. Nojabaei, and P. Khodaparast, "A Data-Based Continuous and Predictive Viscosity Model for the Oil-Surfactant-Brine Microemulsion Phase," in *SPE Improved Oil Recovery Conference?*, 2024, p. D031S019R004.
- [70] Y. Yao, M. Wei, Y. Cui, M. Ali, J. Leng, and Y. Qiu, "Prediction of Surfactant Performance from Surfactant Huff-Puff in Carbonate Reservoirs Using a Data-Driven Approach and Desktop Application Development," in *SPE Western Regional Meeting*, 2024, p. D021S007R005.
- [71] R. S. Alvim, O. A. Babilonia, Y. M. Celaschi, and C. R. Miranda, "Nanoscience applied to oil recovery and mitigation: a multiscale computational approach," *Mrs Advances*, vol. 2, pp. 477-482, 2017.
- [72] A. Gbadamosi, H. Adamu, J. Usman, A. Usman, M. M. Jibril, B. A. Salami, *et al.*, "New-generation machine learning models as prediction tools for modeling interfacial tension of hydrogen-brine system," *International Journal of Hydrogen Energy*, vol. 50, pp. 1326-1337, 2024.
- [73] A. Mohammadi, M. P. Keradeh, A. Keshavarz, and M. Farrokhrouz, "Advanced machine learning-based modeling of interfacial tension in the crude oil-brine-diethyl ether system: insights into the effects of temperature and salinity," *Journal of Molecular Liquids*, vol. 404, p. 124861, 2024.
- [74] M. Maleki, M. R. Dehghani, A. Akbari, Y. Kazemzadeh, and A. Ranjbar, "Investigation of wettability and IFT alteration during hydrogen storage using machine learning," *Heliyon*, vol. 10, 2024.
- [75] E. Salehi, M.-R. Mohammadi, A. Hemmati-Sarapardeh, V. R. Mahdavi, T. Gentzis, B. Liu, *et al.*, "Modeling interfacial tension of N₂/CO₂ mixture+ n-alkanes with machine learning methods: application to eor in conventional and unconventional reservoirs by flue gas injection," *Minerals*, vol. 12, p. 252, 2022.
- [76] C. Liu, J. Wang, J. Wang, and A. Yarahmadi, "Accurate modeling of crude oil and brine interfacial tension via robust machine learning approaches," *Scientific Reports*, vol. 14, p. 28800, 2024.
- [77] G. R. Vakili-Nezhaad, A. Al Ajmi, A. Al Shaaili, F. Mohammadi, and A. Kazemi, "Machine learning assisted modeling of interfacial tension in the system N₂/Brine," *Sustainable Chemistry and Pharmacy*, vol. 33, p. 101071, 2023.
- [78] S. Rosales, K. Zapata, F. B. Cortes, B. Rojano, C. Diaz, C. Cortes, *et al.*, "Simultaneous Detection of Carbon Quantum Dots as Tracers for Interwell Connectivity Evaluation in a Pattern with Two Injection Wells," *Nanomaterials*, vol. 14, p. 789, 2024.
- [79] K. Ali Abd Al-Hameed, "Spearman's correlation coefficient in statistical analysis," *International Journal of Nonlinear Analysis and Applications*, vol. 13, pp. 3249-3255, 2022.
- [80] V. Plevris, G. Solorzano, N. P. Bakas, and M. E. A. Ben Seghier, "Investigation of performance metrics in regression analysis and machine learning-based prediction models," 2022.
- [81] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *Peerj computer science*, vol. 7, p. e623, 2021.
- [82] H. C. Parks, T. M. McCoy, and R. F. Tabor, "Carbon quantum dot assisted adsorption of graphene oxide to the oil-water interface for copper sensing emulsions," *Advanced Materials Interfaces*, vol. 6, p. 1900392, 2019.
- [83] L. Lai, X.-Q. Wei, W.-H. Huang, P. Mei, Z.-H. Ren, and Y. Liu, "Impact of carbon quantum dots on dynamic properties of BSA and BSA/DPPC adsorption layers," *Journal of Colloid and Interface Science*, vol. 506, pp. 245-254, 2017.