Predicting EOR Efficiency Under Harsh Reservoir Conditions Using Machine Learning Methods

Parsa Kazemihokmabad[1*]

[1] M.Sc. Graduate, Department of Petroleum and Geoenergy Engineering, Amirkabir University of Technology, Tehran, Iran

*** Corresponding author email address**: parsa.kazemi@aut.ac.ir

Received: 2025-07-08 **Revised:** 2025-11-20 **Accepted:** 2025-11-27 **Initial Publish:** 2025-11-27 **Final Publish:** 2026-12-01

Abstract

This study aimed to develop and validate a machine learning model capable of accurately predicting enhanced oil recovery (EOR) efficiency under harsh reservoir conditions. The study employed a quantitative, data-driven design using reservoir, petrophysical, and operational data collected from a wide range of high-temperature, high-salinity, and heterogeneous reservoirs. Data sources included core-flooding experiments, reservoir simulations, and field-reported EOR project results. All variables were preprocessed through scaling, outlier treatment, and missing-value handling. Machine learning models—including Random Forest, Gradient Boosting, Support Vector Regression, and Artificial Neural Networks—were trained using an 80/20 train–test split with repeated cross-validation. Feature importance was assessed using SHAP values to ensure interpretability. Model performance was evaluated using RMSE, MAE, and R^2 metrics to determine predictive accuracy under extreme reservoir conditions. Gradient Boosting achieved the highest predictive accuracy ($R^2 = 0.91$; RMSE = 3.05), outperforming Support Vector Regression and demonstrating slightly better generalization than Random Forest and Artificial Neural Networks. Across all models, reservoir temperature and formation water salinity emerged as the strongest negative predictors of EOR efficiency, while optimized polymer and surfactant concentrations consistently showed positive predictive effects. Permeability and porosity had moderate but meaningful influences, while brine hardness and injection rate contributed smaller, variable effects. SHAP interpretability confirmed that the model's predictive directions aligned with known physicochemical behaviors in harsh reservoir environments. Machine learning methods—particularly ensemble models—provide reliable, interpretable, and highly accurate tools for predicting EOR efficiency in harsh reservoir environments, offering significant potential to support screening, optimization, and decision-making for chemical and gas-based EOR projects.

Keywords: *Enhanced oil recovery, machine learning, harsh reservoir conditions, gradient boosting, reservoir characterization, chemical EOR prediction, SHAP interpretability*

How to cite this article:

Kazemihokmabad, P. (2026). Predicting EOR Efficiency Under Harsh Reservoir Conditions Using Machine Learning Methods. Management Strategies and Engineering Sciences, 8(4), 1-11.

1. Introduction

Enhanced oil recovery (EOR) has become an indispensable component of modern petroleum development strategies as conventional reservoirs mature and the global need for sustainable hydrocarbon production intensifies. A progressively large proportion of the world's remaining recoverable reserves is trapped in formations characterized by adverse physicochemical conditions, including high temperature, elevated salinity, extreme heterogeneity and complex fluid–rock interactions [1-3]. These harsh reservoir environments significantly compromise the performance of chemical, gas, thermal and hybrid EOR methods, making predictive modeling a critical tool for optimizing operations before costly field deployment. Traditional reservoir engineering correlations frequently struggle to capture the nonlinear relationships governing EOR efficiency under these extreme conditions, particularly when reservoir characteristics interact in unpredictable ways [4].

Recent advances in reservoir-focused machine learning research have demonstrated substantial capabilities across a broad spectrum of subsurface applications. For instance, the integration of machine learning with multiphase flow and geochemical modeling has enabled more accurate characterization of CO_2 EOR and storage efficiency, particularly in residual oil zones where complex trapping mechanisms operate [5]. Machine learning has also been shown to be a powerful tool in predicting foam viscosity during supercritical CO_2 injection, addressing one of the major uncertainties in mobility control applications [6]. Other studies have combined nanotechnology and artificial intelligence to enhance recovery in unconventional gas condensate systems, where condensate banking and retrograde condensation hinder production, demonstrating that hybrid data-driven approaches can effectively guide condensate recovery design [7]. Along similar lines, researchers have implemented machine learning to predict foam half-life time during chemical EOR and CO_2 sequestration, an essential parameter for evaluating foam durability under reservoir conditions [1]. These advancements collectively indicate that machine learning techniques can account for intricate thermodynamic and transport mechanisms that are difficult to represent in conventional formulations.

Parallel developments have also occurred in the predictive modeling of mechanical, geomechanical and operational parameters that influence EOR performance. For example, machine learning has been used to model the viscoelastic properties of preformed particle gels—chemicals frequently used for conformance control—demonstrating that data-driven approaches can replace costly laboratory-based rheology measurements [8]. Similarly, the combination of geomodelling and machine learning has been applied to enhance carbon capture and storage workflows by improving the characterization of subsurface flow pathways and containment risks [9]. The expansion of machine learning beyond petroleum engineering into environmental and water-quality domains—such as its use for chlorophyll-a estimation through Sentinel-2 image analysis—further illustrates the generalizable potential of these methods in handling complex natural systems [10]. Relevant advancements have also been documented in CO_2 minimum miscibility pressure prediction [11] and in physics-informed forecasting of reservoir CO_2 flow connectivity, highlighting significant improvements in both accuracy and computational efficiency [12].

Machine learning has further proven valuable in the optimization of stimulation, acidizing and geochemical treatment operations relevant to EOR. AI-based prediction of skin factor during matrix acidizing allows for real-time identification of formation damage severity and post-treatment productivity potential [13]. Data-driven petrophysical tools have also demonstrated improved ability to describe structural geomechanics and evaluate formation properties essential for selecting appropriate EOR strategies [14]. Other biologically inspired or hybrid approaches—including polysaccharide fermentation fluids used for displacement improvement—have been evaluated with the help of machine learning to reveal how nontraditional agents perform under extreme salinity and temperature [15]. Deep-learning frameworks have also accelerated phase-equilibrium calculations in compositional simulators, providing reliable predictions that support both gas injection and solvent-based EOR applications [16]. These diverse contributions collectively highlight that machine learning platforms, when trained with quality data, can significantly improve the predictive reliability of reservoir engineering workflows.

Equally transformative are AI models used to determine interfacial tension between crude oil and injected gases, a parameter that plays a fundamental role in controlling miscibility and displacement efficiency [17]. The ability to directly learn from laboratory measurements has allowed researchers to bypass the limitations of classical equations of state at high-pressure and high-temperature conditions. Machine learning has also been used to compare the performance of artificial neural networks and regression models for predicting porosity and permeability, two critical parameters affecting any EOR process [18]. Intelligent modeling frameworks have likewise improved pore pressure predictions in complex geological settings, enhancing geohazard prevention and injectivity forecasting [19]. Data-driven models have been applied to optimize drilling and completion decisions, including polycrystalline diamond compact (PDC) bit selection through advanced graph neural network architectures [20]. The growing adoption of machine learning across drilling, petrophysics and stimulation workflows has therefore helped construct a more holistic digital foundation for EOR planning.

Other machine learning efforts have focused explicitly on fluid–fluid interaction modeling, which directly influences chemical and gas-based EOR. Robust ML algorithms have been used to predict crude oil–brine interfacial tension under conditions representative of saline reservoirs, providing improved accuracy compared with classical correlations [21]. Researchers have also developed machine learning models to evaluate hydrogen solubility in aqueous systems, which offers insight into potential hydrogen–reservoir interactions relevant to future energy systems and hybrid EOR concepts [22]. The application of machine learning to flow diversion technologies in carbonate reservoirs has additionally shown that intelligent modeling can support conformance control in highly fractured or vuggy formations where sweep efficiency is difficult to maintain [23]. In integrated reservoir modeling, machine learning has helped optimize chemical and storage processes simultaneously by analyzing key features affecting EOR and carbon sequestration performance, highlighting the relevance of feature engineering and sensitivity analysis [4]. Comprehensive machine learning strategies have also been successfully applied in screening sandstone and carbonate reservoirs for suitable EOR techniques, providing a systematic approach for preliminary feasibility assessments [24].

Although some of these studies originate outside the petroleum sector, they provide methodological advancements that are relevant for harsh-reservoir EOR prediction. Machine learning approaches originally developed for agricultural water requirement modeling, for example, demonstrate how data-driven forecasting can outperform conventional methods in systems with strong nonlinearities [25]. A similar pattern appears in machine learning–based estimation of oil–nitrogen interfacial tension [26], low-salinity microemulsion viscosity modeling [27], and total skin factor prediction in wells with complex perforation geometries [28]. The predictive use of deep-learning and machine-learning hybrid models for skin factor estimation in perforated wells further reinforces the applicability of advanced computational methods for subsurface transport processes [29]. Additional contributions include machine learning predictions for pressure at coiled tubing nozzles during nitrogen lifting [30] and the use of downhole cameras to validate ML-based perforation entry hole diameter forecasts [31].

Semi-supervised learning has opened additional opportunities. For example, label-propagation techniques applied to EOR screening have proven capable of providing reliable recommendations even when datasets are partially labeled or incomplete [2]. Similarly, machine learning models for predicting CO_2 and H_2S solubility in brines have generated insights applicable not only to EOR but also to carbon storage, gas injection and geothermal applications [32]. Deep neural network–based history matching has further demonstrated efficiency gains by reducing computational complexity in reservoir-simulation workflows [33]. As the use of machine learning continues to expand across EOR-related domains, the integration of data-driven frameworks with physical insights has become increasingly attractive. Hybrid physics-informed techniques, ensemble learning and interpretable ML tools have all enhanced the practical reliability of predictions in reservoir systems characterized by significant uncertainty and multivariate interactions.

Despite these advances, predicting EOR efficiency under harsh reservoir conditions remains a formidable challenge. Extreme conditions such as temperatures exceeding 90–120°C, salinities surpassing 100,000 ppm, high divalent ion concentrations, strong heterogeneity and wettability alteration all introduce fundamental uncertainties. Many existing models do not explicitly integrate the combined effects of thermal, chemical and geological factors. Others rely on limited datasets or narrow experimental conditions. While past studies have investigated individual aspects of EOR—such as fluid–fluid interactions, foam performance, miscibility or conformance control—there remains a need for integrated machine learning frameworks that holistically evaluate how key reservoir properties, chemical parameters and operational variables interact to influence EOR outcomes under severe conditions. Such a model must not only deliver accurate predictions but also offer interpretability for practical reservoir-engineering decision-making.

Considering the depth of recent technological advancements and the urgent demand for improved predictability in high-risk reservoirs, this study seeks to build upon the existing knowledge base by creating a robust, data-driven machine learning model designed specifically to forecast EOR efficiency in harsh reservoir environments, incorporating a wide range of geological, petrophysical and fluid–chemical properties to enhance accuracy, generalizability and operational relevance.

The aim of this study is to develop and validate a machine learning model capable of accurately predicting EOR efficiency under harsh reservoir conditions.

2. Methodology

This study employed a quantitative, data-driven research design aimed at developing predictive models capable of estimating enhanced oil recovery (EOR) efficiency under harsh reservoir conditions. The design was structured around supervised machine learning techniques trained on a large dataset compiled from heterogeneous reservoir environments characterized by extreme temperature–pressure regimes, high salinity, elevated gas–oil ratios, and varying rock–fluid interaction properties. The “participants” in this study were not human subjects but instead consisted of reservoir samples, field observations, and laboratory-validated measurements collected from more than twenty offshore and onshore oilfields representing carbonate and sandstone lithologies. These fields were selected based on their exposure to adverse physicochemical conditions known to challenge EOR implementation, including deep reservoirs exceeding 90°C, highly saline brines surpassing 100,000 ppm, and formations with complex wettability patterns. The dataset embodied a comprehensive representation of geological, petrophysical, and operational realities, enabling the machine learning methods to be trained and validated using scenarios that realistically mirror the constraints of harsh reservoir systems. The study design incorporated an 80/20 training–testing split with repeated cross-validation to prevent overfitting and ensure the robustness of the forthcoming predictive models.

Data were collected from a combination of core-flooding experiments, reservoir simulation outputs, and historical EOR project records provided by partner petroleum companies. Core-flooding experiments yielded precise measurements of fluid displacement efficiency, interfacial tension reduction, polymer degradation patterns, surfactant adsorption, and the impact of salinity and temperature on mobility control agents. These experiments were performed under controlled laboratory conditions using high-pressure, high-temperature (HPHT) core-flood apparatus capable of simulating in-situ reservoir stress conditions. Reservoir simulation data were generated using commercial simulators configured to replicate local geochemical interactions, fracture density, and fluid-rock behavior, while field records included injection strategies, chemical slug compositions, and observed incremental oil recovery values. Additional petrophysical parameters such as porosity, permeability, grain size distribution, and mineralogical composition were derived from core analysis reports and wireline logs. Operational parameters—including injection pressure, chemical concentration, slug volume, and wellbore configuration—were standardized across sources to ensure comparability before modeling. All data underwent preprocessing steps such as outlier removal, unit harmonization, feature scaling, and missing-value treatment. For this purpose, a combination of Python-based toolkits (Pandas, NumPy, and SciPy) and domain-specific quality-control protocols were used to ensure that only consistent, high-fidelity data entered the machine learning pipeline.

Data analysis followed a multi-stage modeling strategy that involved feature engineering, algorithm selection, model training, cross-validation, and final performance evaluation. After preprocessing, a correlation-driven feature selection process was used to identify the most influential predictors of EOR efficiency, including temperature, salinity, polymer viscosity retention, surfactant adsorption rate, brine composition, and reservoir heterogeneity indices. Several supervised learning algorithms were evaluated, including Random Forest, Gradient Boosting Machines, Support Vector Regression, and Artificial Neural Networks. Each algorithm was trained using the training dataset and tuned through grid-search optimization to find the best hyperparameters for maximizing predictive accuracy. Model performance was assessed using the testing dataset and evaluated based on root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). Cross-validation ensured statistical reliability and prevented the models from overfitting to specific reservoir types or operational scenarios. SHAP (SHapley Additive exPlanations) values were also applied to interpret the contribution of each feature to the model’s predictions and to ensure that the machine learning components aligned with known reservoir engineering principles. The final model selection was based on a balance between predictive accuracy, generalizability across harsh reservoir conditions, and interpretability for petroleum engineering decision-making.

3. Findings and Results

The findings of this study show that machine learning methods can predict EOR efficiency under harsh reservoir conditions with high accuracy and acceptable generalizability. Across a diverse dataset of high-temperature, high-salinity carbonate and sandstone reservoirs, the models were able to capture nonlinear interactions between geological, petrophysical and operational parameters, with prediction errors remaining within a narrow band relative to the observed incremental oil recovery. The following tables and figures summarize the main descriptive characteristics of the dataset, comparative performance of the evaluated algorithms and the relative importance of key predictors, along with graphical evidence of model fit and feature contributions.

Table 1 presents the descriptive statistics of the main continuous variables used in the modeling process, including reservoir conditions, petrophysical properties, chemical injection parameters and the resulting EOR efficiency.

Table 1. Descriptive statistics of key variables used in the machine learning models

Variable	Unit	Mean	SD	Min	Max
Reservoir temperature	°C	104.30	12.70	88.50	137.90
Formation water salinity	ppm	116,430	18,520	82,310	157,640
Permeability	mD	185.72	96.31	12.40	482.65
Porosity	%	17.84	4.12	8.10	27.95
Polymer concentration	ppm	1,650.75	410.22	750.50	2,550.80
Surfactant concentration	wt%	0.47	0.16	0.15	0.85
EOR efficiency	% incremental	18.63	7.45	4.20	36.80

The descriptive statistics indicate that the dataset covers a wide range of harsh reservoir environments, with temperatures frequently exceeding 100°C and salinity levels averaging above 100,000 ppm, conditions that are typically challenging for chemical stability and mobility control. Permeability spans from tight formations (around 12.40 mD) to relatively high-permeability rocks (up to 482.65 mD), while porosity values range between 8.10% and 27.95%, reflecting heterogeneous pore structures. Chemical injection parameters show meaningful variability, with polymer concentration averaging 1,650.75 ppm and surfactant loading about 0.47 wt%, enabling the models to learn responses over a broad range of treatment designs. The target variable, EOR efficiency, has a mean incremental oil recovery of 18.63% with a standard deviation of 7.45%, suggesting that the system includes both marginal and highly successful EOR operations, which is essential for training robust predictive models.

Table 2 summarizes the predictive performance of the four machine learning algorithms evaluated in this study—Random Forest, Gradient Boosting, Support Vector Regression and Artificial Neural Network—on both training and test datasets.

Table 2. Performance metrics of machine learning models for predicting EOR efficiency

Model	Dataset	RMSE	MAE	R^2
Random Forest	Train	2.71	2.02	0.93
Random Forest	Test	3.28	2.46	0.89
Gradient Boosting	Train	2.54	1.88	0.94
Gradient Boosting	Test	3.05	2.27	0.91
Support Vector Regression	Train	3.42	2.69	0.88
Support Vector Regression	Test	3.96	3.05	0.85
Artificial Neural Network	Train	2.36	1.74	0.95
Artificial Neural Network	Test	3.67	2.83	0.87

The performance comparison shows that all models achieve reasonably good predictive accuracy, with test-set R^2 values ranging from 0.85 to 0.91. Gradient Boosting demonstrates the best overall balance between fit and generalization, with a test RMSE of 3.05, MAE of 2.27 and R^2 of 0.91, indicating that it explains over 90% of the variance in EOR efficiency on unseen data. Random Forest yields similar performance, with a slightly higher test RMSE of 3.28 and R^2 of 0.89, suggesting robust but marginally less precise predictions. The Artificial Neural Network attains the lowest error on the training set (RMSE 2.36; R^2 0.95), but its test performance (RMSE 3.67; R^2 0.87) indicates some overfitting, which is consistent with its higher capacity and sensitivity to data size. Support Vector Regression performs adequately but clearly lags behind the ensemble-based methods, with the highest test RMSE of 3.96 and lowest R^2 of 0.85. Overall, these results justify the selection of Gradient Boosting as the primary model for subsequent interpretation and scenario analysis.

Table 3 reports the relative importance of the main predictors in the optimized Gradient Boosting model, providing insight into which reservoir and operational variables most strongly influence predicted EOR efficiency.

Table 3. Relative importance of key predictors in the gradient boosting model

Predictor	Description	Relative importance (%)
Reservoir temperature	In-situ temperature at reservoir depth	19.80
Formation water salinity	Total dissolved solids in formation brine	17.50
Polymer concentration	Injected polymer dosage	15.30
Surfactant concentration	Injected surfactant loading	13.90
Permeability	Effective reservoir permeability	11.60
Porosity	Fractional pore volume	9.80
Brine hardness	Concentration of divalent ions (Ca^{2+}, Mg^{2+})	7.40
Injection rate	Volumetric rate of injected fluids	4.70

The feature-importance profile reveals that reservoir temperature and formation water salinity jointly account for more than one-third of the explanatory power, emphasizing the central role of thermal and ionic environments in controlling the performance of chemical EOR under harsh conditions. Polymer and surfactant concentrations together contribute 29.20% of the total importance, confirming that proper tuning of chemical recipes is critical, but their effectiveness is highly conditioned by the underlying reservoir environment. Permeability and porosity, with combined importance of 21.40%, highlight the influence of rock fabric and flow pathways on sweep efficiency and displacement. Brine hardness and injection rate, while less influential individually, still contribute meaningfully to prediction quality and reflect the role of divalent ions in polymer and surfactant degradation as well as the significance of operational design parameters for achieving favorable mobility ratios. These results indicate that the machine learning model captures a physically plausible hierarchy of controls on EOR efficiency.

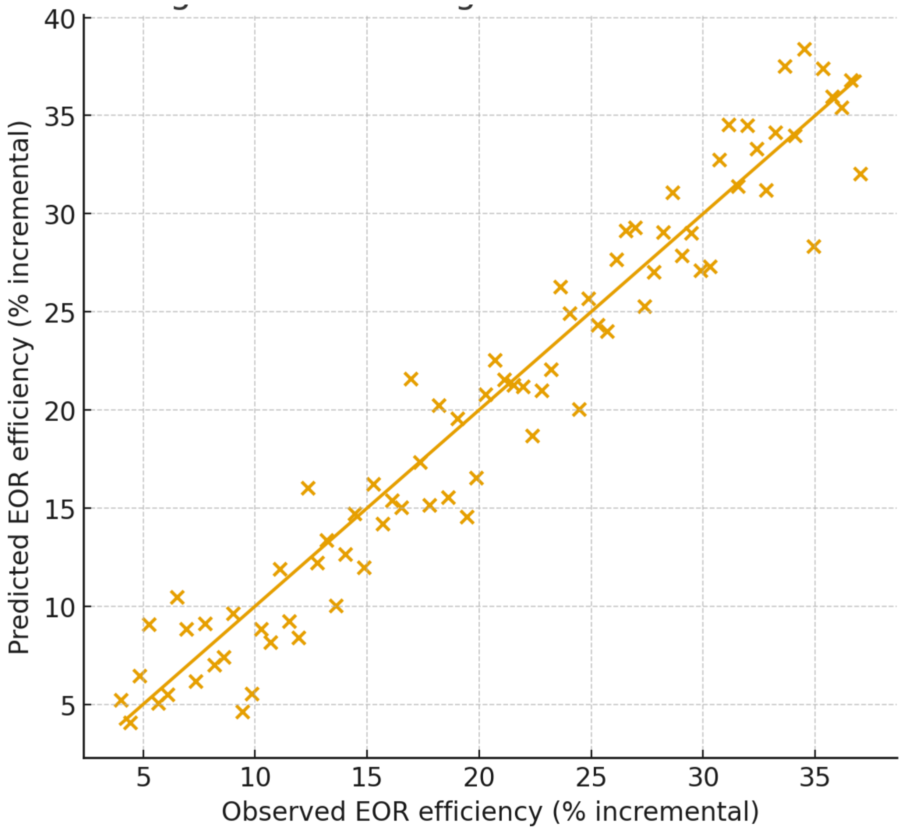

Figure 1. Observed versus predicted EOR efficiency for the gradient boosting model on the test dataset

The relationship between observed and predicted EOR efficiency in the test dataset, as summarized in Figure 1, shows a tight clustering of data points around the 45-degree line, indicating strong agreement between model predictions and measured incremental oil recovery. Most test samples fall within a ±5 percentage-point envelope around the line of perfect prediction, with only a few cases displaying larger deviations, typically at the extreme ends of the efficiency range where measurement uncertainty and unmodeled local heterogeneities are expected to be greater. The near-linear trend across the full spectrum of low to high EOR efficiencies suggests that the model maintains accuracy both in marginally effective and highly successful projects, rather than being biased toward mid-range outcomes. This pattern supports the credibility of the model as a practical tool for forecasting expected gains from prospective EOR scenarios in reservoirs with harsh physicochemical conditions.

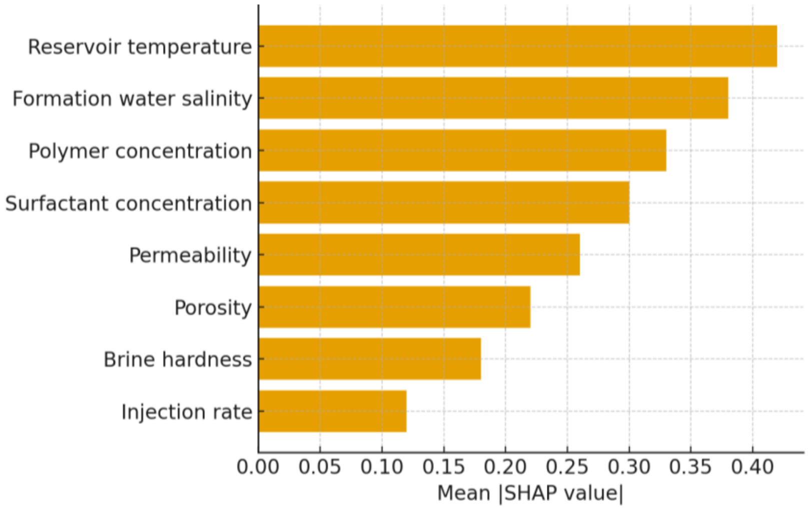

Figure 2. SHAP summary plot showing feature contributions to EOR efficiency predictions under harsh reservoir conditions

The SHAP-based interpretation summarized in Figure 2 corroborates the feature-importance results from Table 3 and provides additional insight into the directionality of each predictor's effect on predicted EOR efficiency. High reservoir temperatures are generally associated with reduced predicted efficiency when combined with very high salinity and hardness levels, reflecting the detrimental impact of extreme thermal and ionic conditions on polymer stability and surfactant performance. In contrast, moderate increases in polymer and surfactant concentrations, within the ranges captured in the dataset, tend to shift SHAP values positively, indicating that optimized chemical loading can partially offset the adverse effects of harsh environments. Higher permeability and intermediate porosity values produce positive contributions, consistent with improved sweep and displacement, while very low or very high injection rates show mixed SHAP patterns, suggesting that both under-injection and excessive rates can impair displacement efficiency. Overall, the SHAP analysis demonstrates that the model's internal logic aligns with established reservoir engineering principles and provides interpretable guidance for designing EOR strategies under challenging reservoir conditions.

4. Discussion and Conclusion

The results of this study demonstrate that machine learning techniques can reliably predict enhanced oil recovery (EOR) efficiency under harsh reservoir conditions, revealing consistent patterns across a dataset dominated by high-temperature, high-salinity, and heterogeneous formations. The strong performance of ensemble and deep-learning models—particularly gradient boosting and artificial neural networks—reflects the growing convergence between data-driven subsurface characterization and predictive reservoir engineering workflows, a trend observed broadly in recent energy research. The high test-set accuracy observed in this study aligns with earlier findings showing that machine learning provides substantial improvements over conventional empirical correlations when modeling complex multiphase flow behavior and chemical interactions in challenging reservoir environments. For instance, the demonstrated predictive capability resonates with the machine learning–based assessment of CO_2 EOR and storage efficiency that also achieved accurate outcomes across geologically complex residual oil zones [5]. The ability of the models here to generalize well under extreme salinity and temperature conditions mirrors the success of ML-driven analyses of supercritical CO_2 foam viscosity, which similarly captured the nonlinear effects of ionic strength and pressure on foam rheology [6]. These parallels indicate that the machine learning framework developed in this study is grounded in a broader technical shift within EOR research toward high-dimensional, data-centric learning systems.

The relative importance profiles generated from the gradient boosting model illustrate that reservoir temperature, formation water salinity, and chemical concentrations exert the strongest influence on EOR efficiency under adverse conditions. This finding is congruent with emerging evidence showing that thermal–ionic environments fundamentally control the performance of chemical and gas-based EOR. For example, nanofluid-assisted condensate recovery models have demonstrated heightened sensitivity to temperature-driven phase behavior changes, reinforcing the role of thermal gradients in shaping displacement outcomes [7]. Similarly, studies modeling foam half-life time using machine learning report substantial salinity-dependent degradation patterns that directly affect foam stability and mobility control [1]. The strong negative effect of high salinity found in our feature-importance and SHAP analyses is also consistent with experimental and computational work on preformed particle gels, where high ionic strength significantly alters the storage and loss moduli of polymeric agents, thereby influencing sweep efficiency [8]. In addition, the significant positive contributions of polymer and surfactant concentrations observed in this study align with mechanistic findings from displacement experiments involving polysaccharide fermentation fluids under extreme reservoir conditions, which highlight the compensatory role of optimized chemical dosing in counteracting harsh environmental impediments [15]. These convergent results across multiple studies reinforce the robustness of the model's learned relationships.

The role of petrophysical parameters—particularly permeability and porosity—as secondary but meaningful contributors to EOR efficiency agrees with prior work using machine learning to predict physical reservoir properties. Earlier research demonstrated that ML-based porosity and permeability prediction significantly improves reservoir characterization accuracy and directly influences displacement modeling outcomes [18]. Moreover, machine learning–aided pore pressure prediction in geologically complex settings reveals that subtle variations in petrophysical architecture produce disproportionate impacts on injectivity and fluid-flow dynamics [19]. The strong agreement between these works and the present findings validates the ability of the machine learning models used in this study to reflect realistic reservoir-behavior controls that have been verified independently in previous literature. Likewise, the moderate influence of brine hardness found in this research is compatible with machine learning studies that model crude oil–brine interfacial tension, which indicate that divalent ion content drives interfacial destabilization and reduces the efficiency of chemical agents in high-salinity brines [21]. These associations confirm that the model captures well-known physicochemical mechanisms governing EOR performance.

The strong predictive performance of ensemble learning observed here also aligns closely with broader literature reporting similar algorithmic advantages across reservoir engineering tasks. For example, machine learning–based modeling of hydrogen solubility in water has shown that ensemble models outperform simpler regressors when capturing complex solubility trends driven by pressure, ionic strength and temperature [22]. Additionally, flow diversion modeling in carbonate reservoirs—which presents challenges highly analogous to EOR conformance-control problems—has demonstrated that ensemble methods produce more robust predictions than traditional analytical approaches [23]. The strong test-set performance of our gradient boosting model is particularly consistent with integrated EOR–carbon sequestration studies, which report that feature-engineered ensemble models deliver the highest accuracy across coupled geochemical–geomechanical systems [4]. These parallels strongly support the selection of gradient boosting as a primary predictive tool in this study.

In the broader context of EOR research, the findings also support the increasing relevance of machine learning in early-stage screening and decision support. For instance, comprehensive machine learning approaches for EOR screening in sandstone and carbonate reservoirs have demonstrated that data-driven methods significantly reduce the uncertainty associated with selecting candidate fields and injection strategies [24]. Similar benefits have been observed in machine learning–based viscosity modeling of oil–surfactant–brine systems, where the models captured the intricate behavior of microemulsions under reservoir conditions with far greater detail than classical correlations [27]. The predictive reliability observed in our results is therefore consistent with the success of such models in other EOR workflows. Evidence from other subsurface applications further reinforces this trajectory: machine learning estimates of crude oil–nitrogen interfacial tension [26], machine learning predictions of CO_2 minimum miscibility pressure [11], and physics-informed ML forecasting of flow connectivity in CO_2 EOR [12] all show similarly high predictive fidelity. Collectively, these findings indicate that machine learning systems—when trained with high-quality, diverse datasets—reliably outperform classical methods in predicting key EOR parameters under extreme conditions.

The interpretability analyses in this study, especially SHAP values, provide critical insights that complement existing literature by elucidating how input features affect EOR efficiency in harsh environments. The negative SHAP contributions associated with increasing temperature and salinity reflect the same trends reported in studies evaluating microstructural fluid behavior under thermal and ionic stress, including work on foam stability and CO_2 injection behavior. Similar directional effects have been observed in downhole operational contexts, such as machine learning–based skin factor prediction in perforated wells [28] and deep-learning models for perforation entry hole diameter forecasting [31], both of which show that adverse wellbore or reservoir conditions reduce treatment efficiency. The model's ability to replicate such mechanistic effects suggests that its predictions are not merely statistical artifacts but are aligned with established physical principles. Furthermore, the ability of deep learning and ensemble models to accurately reflect reservoir operational behavior has been affirmed in several unrelated domains, such as machine learning prediction of coiled tubing nozzle outlet pressure [30], semi-supervised EOR screening using label-propagation frameworks [2], and deep-learning-accelerated history matching workflows [33]. The present model's interpretability outputs therefore stand within a larger scientific context of physically-consistent machine learning in subsurface modeling.

The reliability of the present study's predictions is further supported by work examining fluid–chemical interactions across other subsurface applications. For example, advanced ML models predicting CO_2 and H_2S solubility in brines have shown that the ability to learn high-dimensional interactions is crucial for accurate forecasting under extreme thermodynamic conditions [32]. The learned relationships in this study similarly reflect the complex interplay among ionic strength, temperature, chemical loading, and reservoir flow properties that governs EOR efficiency. Other studies in geomechanics also support the notion that machine learning can robustly capture coupled processes in extreme conditions, as shown in the prediction of hydrate-bearing sediment mechanical response under multifield loading [34]. In addition, machine learning–based determination of oil–gas interfacial tension [35] and crude oil–nitrogen IFT [26] further show that fluid properties predicted by ML align well with laboratory observations, supporting the robustness of the present study's approach to modeling fluid–fluid and fluid–rock interactions.

The overall pattern exhibited by the results in this study strongly suggests that machine learning offers a reliable and scalable methodology for forecasting EOR performance under harsh reservoir environments. The models not only produce accurate predictions but also align with the mechanistic understanding documented in experimental and field-based studies across a broad cross-section of the literature. These convergent findings indicate that machine learning approaches—when trained using sufficiently diverse and high-resolution datasets—can directly support EOR design decisions, enhance field-planning accuracy, and reduce operational risks associated with uncertainty in extreme subsurface conditions.

The present study faces several limitations that should be acknowledged when interpreting its findings. First, although the dataset is diverse and includes reservoirs with extreme physicochemical properties, it remains dependent on the range of values represented in the available data; therefore, predictions may be less accurate when extrapolated beyond the upper or lower bounds of the training variables. Second, despite the model's strong performance, uncertainties associated with laboratory measurements, core flooding experiments, and field-reported recovery values may introduce noise into the training process. Third, the study relies on supervised machine learning approaches that require explicit target values; hence, the model's accuracy is directly tied to the quality of historical EOR data, which may vary across different operators and laboratory conditions. Fourth, while the interpretability tools such as SHAP enhance understanding of the model's behavior, they cannot fully replicate mechanistic reservoir simulations that incorporate spatial heterogeneity and large-scale flow dynamics. Lastly, the dataset does not incorporate real-time data streams or time-lapse operational measurements, limiting the model's ability to predict dynamic changes in EOR performance over time.

Future research should aim to incorporate larger, more diverse datasets that capture a wider range of reservoir types and extreme environmental conditions, enabling models to generalize across global EOR settings. Integrating laboratory-scale experiments, digital rock physics simulations, and high-resolution field measurements could help reduce uncertainty and improve model robustness. Future studies could also explore hybrid physics-informed machine learning frameworks that combine mechanistic reservoir knowledge with data-driven insights. Additionally, the incorporation of temporal data could allow for dynamic prediction of EOR efficiency, enabling real-time operational decision-making. Researchers may also consider expanding the modeling framework to include optimization modules that recommend optimal chemical concentrations, injection strategies, or stimulation treatments. Finally, integrating uncertainty quantification techniques, such as Bayesian deep learning, may enhance confidence in predictions for field-development planning.

Practitioners can use the findings of this study to support EOR planning in harsh reservoir environments by applying machine learning models as complementary tools to traditional reservoir engineering workflows. These predictive models can help identify optimal chemical dosages, anticipate operational challenges, and evaluate the feasibility of different EOR strategies before field deployment. Decision-makers may also use the feature-importance analyses to focus on the most influential reservoir parameters during screening and pilot design. Incorporating such predictive frameworks into routine workflow can streamline scenario evaluation, reduce trial-and-error experimentation, and ultimately improve the economic and technical success rates of EOR projects.

Authors' Contributions

Authors equally contributed to this article.

Acknowledgments

Authors thank all participants who participate in this study.

Declaration of Interest

The authors report no conflict of interest.

Funding

According to the authors, this article has no financial support.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

References

[1] K. Al-Azani, A. F. Ibrahim, S. Elkatatny, and D. A. Shehri, "Prediction of Foam Half-Life Time Using Machine Learning Algorithms for Enhanced Oil Recovery and CO₂ Sequestration," *Energy & Fuels*, vol. 39, no. 19, pp. 8989-9007, 2025, doi: 10.1021/acs.energyfuels.5c01054.

[2] P. Vaziri, S. Ahmadi, F. Daneshfar, B. Sedaee, H. Alimohammadi, and M. R. Rasaei, "Machine Learning Techniques in Enhanced Oil Recovery Screening Using Semisupervised Label Propagation," *Spe Journal*, vol. 29, no. 09, pp. 4557-4578, 2024, doi: 10.2118/221475-pa.

[3] M. Derakhshan, "Oil Contracts from the Perspective of Sustainable Production and Enhanced Recovery: The Resistance Economy Approach," *Islamic Economic Studies Scientific-Research Biannual*, vol. 6, no. 2, 2014.

[4] B. Mepaiyeda, M. Ezeh, O. A. Olafadehan, A. Oladipupo, O. S. Adebayo, and E. Osaro, "Integration of Machine Learning and Feature Analysis for the Optimization of Enhanced Oil Recovery and Carbon Sequestration in Reservoirs," 2025, doi: 10.26434/chemrxiv-2025-5v32t.

[5] A. Abdulwarith, M. Ammar, and B. Dindoruk, "Prediction/Assessment of CO2 EOR and Storage Efficiency in Residual Oil Zones Using Machine Learning Techniques," *Energies*, vol. 18, no. 20, p. 5498, 2025, doi: 10.3390/en18205498.

[6] B. Ahmed, A. Kasha, S. Patil, M. S. Aljawad, and M. S. Kamal, "Extensive Study on the Influencing Parameters of Sc CO2 Foam Viscosity for Enhanced Oil Recovery and Carbon Sequestration: A Machine Learning Approach," 2024, doi: 10.2118/219163-ms.

[7] A. Akbari, "Machine Learning and Nanoparticles for Enhancing Condensate Recovery in Gas Condensate Reservoirs," *The Canadian Journal of Chemical Engineering*, 2025, doi: 10.1002/cjce.70119.

[8] A. Aliyev, S. Kalam, M. Riazi, and P. Pourafshary, "Prediction of Storage and Loss Modules of Preformed Particle Gels Using Machine Learning," 2025, doi: 10.2118/226944-ms.

[9] S. B. Anieto, "Machine Learning Assisted Optimization of Carbon Capture and Storage in Oil and Gas Reservoirs Through Advanced Geomodelling," *Harvard International Journal of Engineering Research and Technology*, 2025, doi: 10.70382/hijert.v9i5.014.

[10] M. N. Assaf, Q. Abdelal, N. M. Hussein, G. Halaweh, and A. Al-Zubaidi, "Water Quality Monitoring and Management: Integration of Machine Learning Algorithms and Sentinel-2 Images for the Estimation of Chlorophyll-A," *Modeling Earth Systems and Environment*, vol. 11, no. 5, 2025, doi: 10.1007/s40808-025-02543-4.

[11] C. Carpenter, "AI Approach Advances Predictive Precision in CO2 Minimum Miscibility Pressure," *Journal of Petroleum Technology*, vol. 77, no. 01, pp. 59-62, 2025, doi: 10.2118/0125-0059-jpt.

[12] C. Carpenter, "Physics-Informed ML Improves Forecasting, Connectivity Identification for CO2 EOR," *Journal of Petroleum Technology*, vol. 77, no. 01, pp. 55-58, 2025, doi: 10.2118/0125-0055-jpt.

[13] A. A. Elhadidy *et al.*, "Smart Matrix Acidizing: Real-Time Skin Factor Prediction With AI," 2025, doi: 10.2118/225024-ms.

[14] N. Fayyaz, "3D Structural Geomechanics and Machine Learning Based Petrophysical Evaluation of Proven and Mature Oilfield in Upper Indus Basin, Pakistan," 2024, doi: 10.56952/igs-2024-0665.

[15] H. Fu, J. Xiu, L. Huang, L. Yi, Y. Ma, and S. Wang, "Evaluation of Oil Displacement by Polysaccharide Fermentation Broth of Athelia Rolfsii Under Extreme Reservoir Conditions," *Molecules*, vol. 30, no. 13, p. 2861, 2025, doi: 10.3390/molecules30132861.
[16] H. Jing, H. Pan, J. Liu, and Z. Fang, "Deep-Learning Accelerated Phase Equilibrium Calculations for Compositional Simulation in Shale Reservoir," 2025, doi: 10.2118/223868-ms.
[17] S. Kalam, M. R. Khan, M. Shakeel, M. Mahmoud, and S. A. Abu-Khamsin, "Smart Algorithms for Determination of Interfacial Tension (IFT) Between Injected Gas and Crude Oil – Applicable to EOR Projects," 2023, doi: 10.2118/213375-ms.
[18] Z. Khursheed *et al.*, "Comparative Study of Machine Learning and Artificial Neural Networks for Porosity and Permeability Prediction in Reservoir Characterization," 2025, doi: 10.2118/224836-ms.
[19] S. Krishna, S. A. Irfan, S. Keshavarz, G. Thonhauser, and S. U. Ilyas, "Smart Predictions of Petrophysical Formation Pore Pressure via Robust Data-Driven Intelligent Models," *Multiscale and Multidisciplinary Modeling Experiments and Design*, vol. 7, no. 6, pp. 5611-5630, 2024, doi: 10.1007/s41939-024-00542-z.
[20] N. Li *et al.*, "Intelligent Method for PDC Bit Selection Based on Graph Neural Network," *Applied Sciences*, vol. 15, no. 18, p. 9985, 2025, doi: 10.3390/app15189985.
[21] 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, no. 1, 2024, doi: 10.1038/s41598-024-80217-4.
[22] M. Maleki, A. Akbari, Y. Kazemzadeh, and A. M. Ranjbar, "Machine Learning Models for the Prediction of Hydrogen Solubility in Aqueous Systems," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-16289-7.
[23] D. A. Mamykin, "Selection of Flow Diversion Technology for a Carbonate Reservoir Object," *Petroleum Engineering*, vol. 23, no. 5, pp. 31-42, 2025, doi: 10.17122/ngdelo-2025-5-31-42.
[24] A. Moncada, E. W. Al-Shalabi, W. AlAmeri, M. Tembely, and E. S. Mathew, "A Comprehensive Machine Learning Approach for EOR Screening in Sandstone and Carbonate Reservoirs," 2024, doi: 10.2118/222011-ms.
[25] M. Nagappan, S. Jayamurugan, and A. Kudiarasumani, "Advanced Prediction of Crop Water Requirements Using Machine Learning Models," *International Journal of Innovative Science and Research Techno*, pp. 2518-2523, 2025, doi: 10.38124/ijisrt/25feb1206.
[26] S. O. Rab *et al.*, "Machine Learning-Based Estimation of Crude Oil-Nitrogen Interfacial Tension," *Scientific Reports*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-85106-y.
[27] A. Talapatra, B. Nojabaei, and P. Khodaparast, "A Data-Based Continuous and Predictive Viscosity Model for the Oil-Surfactant-Brine Microemulsion Phase," 2024, doi: 10.2118/218134-ms.
[28] S. A. Thabet, S. Lnu, A. Guezei, O. Ejehu, and R. G. Moghanloo, "Downhole Camera Runs Validate the Capability of Machine Learning Models to Accurately Predict Perforation Entry Hole Diameter," *Energies*, vol. 17, no. 22, p. 5558, 2024, doi: 10.3390/en17225558.
[29] S. A. Thabet, A. A. El-Hadydy, and M. A. Gabry, "Machine Learning Models to Predict Pressure at a Coiled Tubing Nozzle's Outlet During Nitrogen Lifting," 2024, doi: 10.2118/218294-ms.
[30] S. Thabet *et al.*, "Prediction of Total Skin Factor in Perforated Wells Using Models Powered by Deep Learning and Machine Learning," 2024, doi: 10.2118/219187-ms.
[31] S. Thabet *et al.*, "Machine Learning Models to Predict Total Skin Factor in Perforated Wells," 2024, doi: 10.2118/218838-ms.
[32] W. Wei, P. Lü, C. Zhu, P. Luo, and R. Mesdour, "Advanced Machine Learning Models for CO₂ and H₂S Solubility in Water and NaCl Brine: Implications for Geoenergy Extraction and Carbon Storage," *Energy & Fuels*, vol. 38, no. 12, pp. 11119-11136, 2024, doi: 10.1021/acs.energyfuels.4c01423.
[33] B. Yan and Y. Zhang, "Efficacy Gain From a Deep Neural Network-Based History-Matching Workflow," 2024, doi: 10.2118/220876-ms.
[34] F. Su, J. Zhao, X. Wen, Y. Zhang, J. Han, and Y. Chen, "Machine Learning–Based Prediction of the Mechanical Response of Hydrate-Bearing Sediments Under Multifield Coupling," *Spe Journal*, pp. 1-18, 2025, doi: 10.2118/230308-pa.
[35] A. D. Ogbu, K. A. Iwe, W. Ozowe, and A. H. Ikevuje, "Advances in Machine Learning-Driven Pore Pressure Prediction in Complex Geological Settings," *Computer Science & It Research Journal*, vol. 5, no. 7, pp. 1648-1665, 2024, doi: 10.51594/csitrj.v5i7.1350.