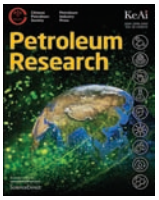
Petroleum Research 11 (2026) 827–837

Contents lists available at ScienceDirect

Petroleum Research

journal homepage: http://www.keaipublishing.com/en/journals/petroleum-research/



Full Length Article

Advanced machine learning approaches for optimizing waterflooding: A focus on injector-producer dynamics



Samuel Nashed *, Oluchi Ejehu, Rouzbeh Moghanloo

Mewbourne School of Petroleum and Geological Engineering, Mewbourne College of Earth and Energy, The University of Oklahoma, Norman, OK 73019, USA

A R T I C L E I N F O

Article history:
Received 7 May 2025
Received in revised form
14 December 2025
Accepted 6 January 2026
Available online 9 January 2026

Keywords:
Waterflooding
Production forecasting
Well connectivity
Machine learning
Enhanced recovery

A B S T R A C T

Waterflooding persists as the primary secondary oil recovery method, although its efficiency evaluation is often hindered by limitations from reservoir heterogeneities alongside non-uniform sweep efficiency and uncertainty in predicting well-to-well connectivity. To improve the evaluation of waterflooding, this research develops and validates nine advanced machine learning models, covering key types such as neural networks, linear models, tree-based models, ensemble learning, and support vector machines. This research utilized a dataset of 6592 observations collected across four wells over 2000 days to train and evaluate these machine learning models for fluid production rate forecasting and injector-producer connectivity assessments. The most successful model was AdaBoost since it delivered both a minimum root mean squared error (RMSE) of 0.004 and a coefficient of determination (R^2) of 0.99, surpassing traditional methods. Furthermore, SHapley Additive Explanations (SHAP) analysis provided interpretability by quantifying injection wells' relative influence on production rates, facilitating the construction of a connectivity-driven contour map. Traditional evaluation methods of empirical and analytical approaches, as well as capacitance-resistance models and numerical simulators, function under restrictive assumptions, extensive computational requirements, and significant data demands. This study is novel in uniting multiple ML algorithms with SHAP interpretability and a connectivity-driven contour map for injector–producer dynamics, providing a practical, real-time workflow for field-scale waterflood optimization. The results of this study demonstrate that ML models surpass traditional methods in terms of predictive accuracy, computer processing speed, and scalability when applied for real-time waterflood optimization. This study illustrates the transformative potential of ML in reservoir engineering by enabling data-driven decision-making and improving the injection optimization process.

© 2026 Chinese Petroleum Society. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Water flooding operates as an essential method across the petroleum industry because it stands among the most applied secondary oil recovery techniques to sustain production and recover additional hydrocarbons from reservoirs (Salehi et al., 2014; Yuliansyah et al., 2021). Natural reservoir drive mechanisms in primary recovery, such as solution gas drive, gas cap drive, gravity drainage, or aquifer support, prove insufficient for mature oil fields, thus resulting in production rate reductions (Mokheimer et al., 2019; Ummah and Setiati, 2022). Typically, the percentage of extractable oil from primary recovery ranges from 5% to 30% and leaves back substantial amounts of oil in place (OOIP) (Yu et al., 2015). To address this challenge, water flooding serves as an essential secondary recovery technique for managing reservoir pressure while boosting the displacement of oil (Xue et al., 2023). Water injection through targeted wells helps to stabilize reservoir pressure and enables efficient water displacement of oil from injection wells towards the producing wells (Wang et al., 2023). Water flooding represents a highly preferred recovery method due to its operational simplicity and economic feasibility when used for conventional reservoirs (Shirazi et al., 2024). Moreover, water flooding stands out against higher complexity enhanced oil recovery (EOR) methods such as gas injection, steam injection, or chemical flooding because it requires lower costs with simple execution (Niedermaier Custodio and Carneiro, 2024).

* Corresponding author.
E-mail addresses: samuel.a.nashed-1@ou.edu (S. Nashed), oluchi.ejehu@ou.edu (O. Ejehu), rouzbeh.gm@ou.edu (R. Moghanloo).
Peer review under the responsibility of Editorial Office of Petroleum Research.

https://doi.org/10.1016/j.ptlrs.2026.01.001
2096-2495/© 2026 Chinese Petroleum Society. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Nomenclature

AdaBoost	Adaptive boosting
CRM	Capacitance resistance model
DCA	Decline curve analysis
DT	Decision trees
EOR	Enhanced oil recovery
GB	Gradient boosting
KNN	K-Nearest Neighbor
L-BFGS	Limited-memory - Broyden - Fletcher - Goldfarb - Shanno optimization algorithm
LR	Linear regression
MAE	Mean absolute error
MAPE	Mean absolute percent error
ML	Machine learning
MSE	Mean square error
NN	Neural network
OOIP	Original oil in place
R^2	Correlation coefficients
RF	Random Forest
RMSE	Root mean square error
SGD	Stochastic gradient descent
SHAP	Shapley additive explanations
SVMs	Support vector machines
WOR	Water - oil ratio

1.1. The importance of waterflood performance evaluation

The assessment of water flooding performance is crucial for achieving optimal oil recovery and making development plans economically viable (Arab et al., 2022). The extensive use of water flooding does not ensure complete oil displacement because reservoir heterogeneities combine with permeability variations and unfavorable mobility ratios (Helmy et al., 2024; Wang et al., 2022). Performance assessment is essential because insufficient monitoring allows water to bypass vital oil reserves, which causes both suboptimal outcomes and early water breakthrough (He et al., 2021). Performance evaluation through proper engineering practices allows engineers to track injection efficiency directly and understand well connectivity, which enables optimized operational adjustments to enhance sweep efficiency and recover more oil (Ding et al., 2021). A thorough understanding of water injection responses from production wells enables engineers to detect performance issues through channeling and conformance problems so they can apply treatment measures including rate adjustments, pattern alignment changes, and conformance control solutions (Li et al., 2021). Effective water flooding operations heavily depend on the proper connection between injection wells and production wells (Patidar et al., 2022). Water flooding uses injection wells to move displaced oil toward production wells, so the strength of well connectivity determines both recovery efficiency and oil sweep effectiveness as well as final reservoir results (Muggeridge et al., 2014). The ultimate goal of water injection should be to create equal distribution of fluid flow throughout the reservoir space to drive oil accumulation to the production zone (Ye et al., 2025). However, the unpredictable movement of injected water occurs because of reservoir heterogeneities, including permeability variations and high-permeability channels, together with natural fractures and faults (Huang et al., 2021; Mohamed et al., 2024b). These conditions produce inefficient displacement and early water breakthrough (Yang and Urdaneta, 2016). Therefore, analyzing injector-producer connectivity enables more efficient optimization of water injection operations, improves sweep efficiency, and reduces water production problems (Castillo Gamarra et al., 2019).

1.2. Traditional evaluation methods

The traditional evaluation methods for water flooding performance deliver essential information about water movement patterns as well as reservoir sweep efficiency alongside the injection response (Shi et al., 2021). Through these methods, engineers gain the ability to measure injection effectiveness in addition to determining forecasted production patterns and maximizing reservoir operation efficiency (Wang et al., 2021). Material balance analysis functions as the basic tool engineers use to understand the complete behavior of reservoirs during fluid displacement periods (Aleev et al., 2017). This technique helps reservoir engineers determine future pressure support and assess reservoir connectivity and potential recovery rates through volumetric calculations and the recorded historical production data. Similarly, DCA is commonly used for forecasting upcoming production patterns through its analysis of the recorded production rate data that fits different types of decline models, including exponential, hyperbolic, and harmonic decline (Batycky et al., 2017). These methods work effectively for reservoir performance evaluation but do not successfully represent complex flow dynamics in heterogeneous reservoirs (Li and Horne, 2003).

Other traditional methods include the water-oil ratio (WOR) analysis, which remains a vital traditional technique for tracking water flood performance through water production pattern recognition, early water breakthrough detection, and high-permeability channel identification (Villamizar et al., 2013). In addition, the Buckley-Leverett theory enables fractional flow and streamline analysis to provide detailed insights about how water injection displaces oil through porous media for timing breakthroughs and enhancing injection strategy optimization (Sumnu-Dindoruk and Dindoruk, 2006). Moreover, tracer tests perform as direct diagnostics through the monitoring of fluid movements between injection wells and production wells, which reveals both formation heterogeneity patterns and well interconnectivity characteristics (Davidescu et al., 2021). However, these evaluation methods inherit their main limitation from their assumption of uniform reservoir characteristics and the need to use outdated production data instead of dynamic interaction monitoring (Haryanto et al., 2019). Most of these established procedures fail to reveal explicit understanding about well-to-well connectivity and do not handle complex reservoir heterogeneities that impact breakthrough times during waterflooding operations (Dandan et al., 2015).

The introduction of Capacitance-Resistance Models (CRMs) serves as a semi-analytical approach that tries to address empirical method limitations for water flooding performance evaluation (Zhang et al., 2018). CRMs derive information about well connectivity and reservoir responses by monitoring the collected rate fluctuations of injection wells and production wells without requiring knowledge of permeability or porosity (Alghamdi et al., 2021). The reservoir appears as a linked system of holding chambers when modeling it with CRMs, where fluid storage amounts correspond to capacitance measures and flow barrier characteristics correspond to resistance values (Sayarpour et al., 2009). CRMs enable the analysis of production well responses to injection

modifications, which leads to vital understanding regarding injector-producer interconnection, time delays, and injection performance metrics (Carpenter, 2020). CRMs demonstrate important advantages over traditional empirical methods since they use only rate data from injection and production wells to determine interwell connectivity (Koroglu et al., 2024). However, CRMs face challenges when analyzing changes to active producer numbers and well productivity indices along with heterogeneous formations, high fluid compressibility variations, and aquifer support presence (Sayarpour et al., 2010). Operating CRMs in large fields with hundreds of wells becomes limited by the excessive computational requirements (Davudov et al., 2020). Additionally, these predictive models lack the capability to effectively account for changing reservoir parameters that are caused by dissolution together with deposition chemical reactions (Temizel et al., 2018). Another drawback is that the models make the inaccurate assumption that the relationship between injection rates and production rates remains linear across all reservoirs, although non-linear interactions could potentially exist (Temizel et al., 2019). These simplified models and restricted assumptions generate inaccurate predictions regarding waterflooding performance during changing reservoir circumstances (Simha et al., 2021).

Another traditional method is numerical reservoir simulation, which stands as a comprehensive approach offering the most complete system for waterflood assessment by building mathematical models that solve fluid flow equations in porous systems (Lavie et al., 2011). These simulators generate digital simulations that replicate reservoir environments to forecast movements of fluids together with pressure variations and production behavior until the end of the simulation period (Li et al., 2013). With reservoir simulation, engineers can analyze heterogeneous formations by building simulation models that combine different geological layers that represent diverse flow units along with comprehensive fluid behavior characteristics (Zeinijahromi et al., 2014). Moreover, numerical reservoir simulation offers solutions for complicated reservoir shapes and complete models of connections among wells (Maucec et al., 2019). However, the implementation of these simulators demands large amounts of data input together with significant computational capabilities, consistent model calibration, and takes considerable processing time and professional knowledge for proper interpretation (Noirot et al., 2003). Additionally, simulation prediction accuracy is significantly influenced by the quality of input data since imperfect reservoir characterization during the modeling process can produce wrong results even though the mathematical models show sophistication (Rao, 2024).

1.3. Machine learning models for waterflood performance evaluation

Machine learning (ML) models attract the oil and gas industry because they solve problems that traditional methods struggle to address (Gharieb et al., 2024c; Nashed and Moghanloo, 2025). Additionally, the increasing availability of processing power together with analysis data makes the application of ML algorithms an attractive prospect for industry stakeholders (Mohamed et al., 2024a; Nashed et al., 2024; Thabet et al., 2024a). Modern waterflood performance evaluation made a breakthrough with ML models because they introduce progressive solutions that transform traditional evaluation methods (Gabry et al., 2025; Koray et al., 2023; Sinha et al., 2024). New industry implementations show how ML models work with massive existing production and injection data to discover advanced connections between wells that traditional methods would miss (Kubota and Reinert, 2019; Li et al., 2024). Recent works also expand the scope of ML in reservoir engineering, including deep learning-based pore network generation for fluid flow analysis (Guo et al., 2024), and ML-driven reservoir permeability prediction (Soromotin et al., 2025), further demonstrating the versatility of data-driven approaches. Unlike analytical and empirical methods, the non-linear characteristics of reservoir behavior together with changes in field conditions can be captured through ML algorithms that operate without needing detailed geological characterization or simplistic assumptions (Hamady et al., 2024; Thabet et al., 2024b). Moreover, the ML models maintain high accuracy levels and reach computational speeds that produce results within minutes instead of the regular weeks dedicated to traditional numerical simulations (Gharieb et al., 2024a; Thabet et al., 2024c). ML models succeed in uniting multiple datasets, such as production data and well test results, to generate complete reservoir knowledge without needing extensive input required by full physics simulators (Asante et al., 2023; Gharieb et al., 2024b). Neural networks, together with ensemble models, demonstrate better accuracy than capacitance-resistance models and need both less calibration work and fewer behavioral assumptions about the reservoir (Kumar et al., 2022). Furthermore, ML models help improve well connectivity assessments and identify problems with water channeling and poor sweep efficiency during the early stages (Madhumaya and Vyas, 2024). In addition, the integration of ML models with the advanced feature-based interpretability method SHAP (SHapley Additive exPlanations) analysis enables efficient evaluation of operational scenarios that include injection redistribution and pattern balancing along with well conversion strategies. Perhaps most importantly, these ML-driven techniques enable operators to optimize waterflood performance across entire fields instead of relying on well or pattern analyses, which results in better reservoir management with reduced computational and personnel demands.

The application of advanced machine learning approaches to predict oil well production rates influenced by water injection and assess the level of communication between injectors and producers remains absent in existing waterflooding performance evaluation studies according to our current knowledge. Although existing literature has used ML to determine interwell connectivity, these studies tend to focus on few algorithms or are not interpretable. Our work is innovative in its systematic benchmark of a wide range of different ML models on the same dataset and incorporation of SHAP analysis to not only measure the effects of the injector-producer but also predict connectivity-based contour maps, therefore, offering a clear and deployable-in-the-field workflow. We believe that the application of ML models presents itself as a precise and efficient alternative that combines cost-effectiveness with easy implementation for waterflooding performance evaluation instead of conventional empirical, analytical, capacitance resistance, and numerical models.

2. Methodology

In this study, the presented methodological framework stretches across four stages in Fig. 1 with specific connections between the phases. The sequential research stages guide the research objectives to develop a coherent alignment.

2.1. Data collection

Measuring both the injection water rate along with the total fluid production rate from the injector and production wells is a common practice and enables optimization of waterflood performance. Our research utilized a comprehensive dataset of 6592

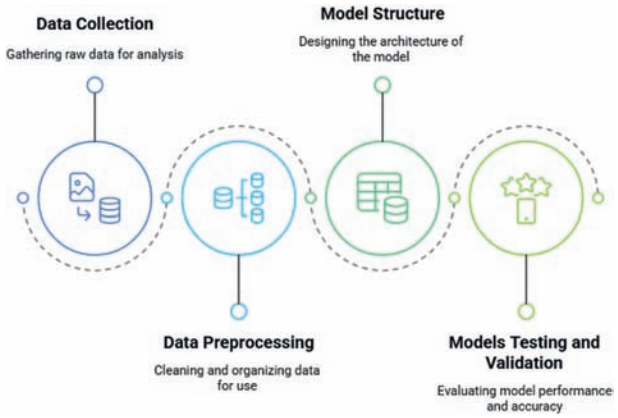


Fig. 1. A diagram of the methodology.

observations obtained from individual measurement data, which served as the basis for this study. A synthetic field with four wells serves as the source of the gathered dataset. The present dataset follows the studies performed by Albertoni and Lake (2003), Albertoni and Lake (2003), Yousef et al. (2006) and Liang et al. (2007). Through this thorough dataset, machine learning models acquire a better understanding of genuine relationships instead of accidental noise patterns during their development phase. Moreover, the examination of large datasets stops overfitting and enables the models to connect complex data patterns. In addition, through complete datasets, machine learning models both optimize their ability to confront realistic conditions and eliminate bias because they display precise representations of all potential situations. Furthermore, the dataset allows appropriate training and testing separation for establishing trustworthy model assessments. The research data includes measurements of injection rate for wells I1, I2, and I3 alongside the total fluid production rate for producer well P4. Figs. 2 and 3 show the injection and production rate of all wells over 2000 days.

2.2. Data preprocessing

The groundwork of any machine learning project is data preprocessing, which describes organized techniques applied to raw data to produce ready-to-use input for future algorithmic processing. Fundamentally, data preprocessing is the entire set of procedures applied to raw data prior to model training, as it operates as a fundamental connection between information collection and pattern discovery. The performance of data processing eliminates unimportant information and noise, which protects models against drawing incorrect conclusions. Data preprocessing follows a planned workflow pattern that begins with cleaning operations that handle missing data and search for abnormal points. The approach that we used to handle missing values included removing complete measurement records and using linear extrapolation from neighboring points to estimate the gaps. In addition, we used box plot statistics for abnormal data point detection, which led to their elimination from the study. This initial work transitions into data integration that unites various formats. Then, standardized and normalized variables create a common measurement scale for strategic data-splitting functions that support training and validation goals.

Machine learning preprocessing depends heavily on data normalization to develop coherent characteristics from features that differ in units and scales for balanced variable contribution to models. The procedure is essential for machine learning algorithms, especially those dependent on distance metrics or gradients, because it produces improved results from features that have standardized scales. Research has validated that normalization practices deliver multiple benefits, which consist of accuracy elevation and faster convergence alongside enhanced computational speed in various machine learning algorithms. The normalization process includes two standard approaches, which are min-max scaling and z-score standardization. Through min-max scaling, the data gets adjusted toward specified ranges without altering original value relationships. The normalization through z-score standardization shifts data to a mean value of zero while achieving unity in standard deviation for statistical purposes. The data normalization techniques have distinct uses that depend on the algorithm's needs alongside the data attributes. This research adopted the Min-Max Scalar normalization technique as its method. Min-max scaling delivers superior benefits for our machine learning needs compared to z-score standardization. The distribution shape remains unchanged by min-max scaling because it only transforms the scale, yet z-score standardization transforms relative distances, which affects data-sensitive algorithms. Additionally, min-max normalization provides support vector machines with exceptional performance because it optimizes the hyperplane in the transformed feature space. The data transformation occurs according to the following equation:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where: X is the original data point, X_{min} is the minimum value in the feature, X_{max} is the maximum value in the feature, X_{scaled} is the normalized value within the [0,1] range.

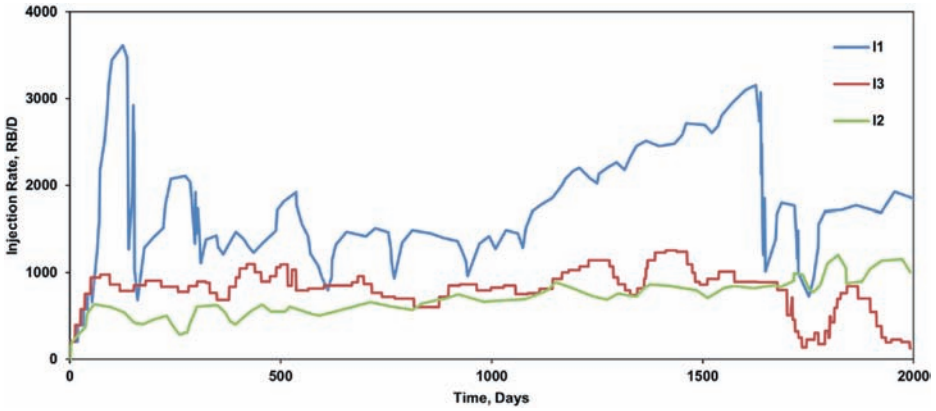


Fig. 2. Individual well injection rate.

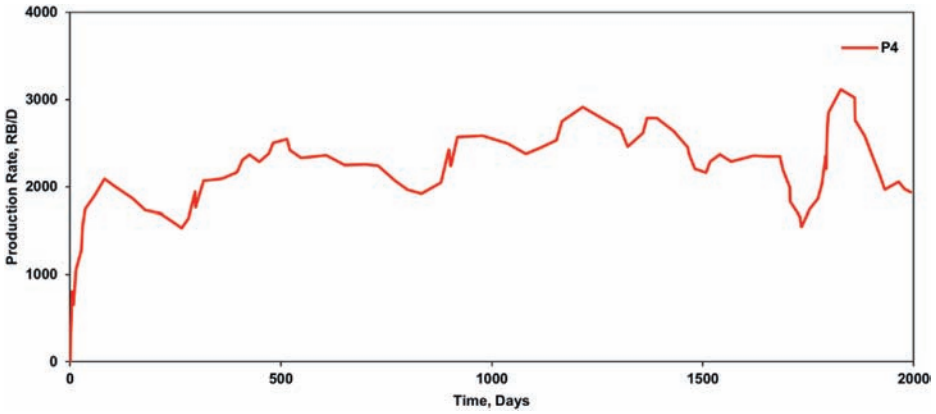


Fig. 3. Well, P4 total fluid production rate.

And at the end, the preprocessed data was divided (80% for training and 20% for testing). We set this ratio because it is a perfect balance between two opposing requirements in the development of machine learning models. 80% of data, which is the finest part, is given to the model for training, where it learns the patterns of data, and the remaining 20% is used for testing, as it is a sufficient amount to judge the performance of the model on unseen examples. This traditional method ensures that model training is reliable and provides statistically meaningful test samples for assessing the ability of generalization, reduces the risk of overfitting, and ensures that the model will operate effectively on the real-world data beyond the training data. Moreover, the 80/20 ratio worked the best of all the ratios in our research using cross-validation and hyperparameter tuning.

2.3. Model structure

The development and evaluation of nine models with Python 3.10.12 required a specific set of hyperparameter settings as outlined in Table 1. Running a multiple regression model such as Gradient Boosting (GB), Adaptive Boosting (AB), Random Forest (RF), Support Vector Machine (SVMs), Decision Tree (DT), K-Nearest Neighbors (KNN), Linear Regression (LR), Neural Network (NN), and Stochastic Gradient Descent (SGD) on the same dataset is a crucial part of this research. This practice allows us to do extensive comparative analysis and figure out which ones work best for our dataset since each model has its strengths. Each machine learning model has its own advantages that are suitable for specific kinds of analytics. For example, Gradient Boosting succeeds in offering very high predictive accuracy via a process of successive error correction, mainly in cases where there is a very high dimensional feature space and complex relationships, while Random Forest wins it in terms of generalization, by virtue of an ensemble of decision trees with the ability of feature importance estimation, and seems to perform better in the high dimensionality data. While Support Vector Machines have the great ability to define optimal decision boundaries by kernel transformations, in comparison to the easy understanding and clear decision-making ability of Decision Trees. K-Nearest Neighbors offers simplicity and effortless training. Linear regression is able to keep the mathematical interpretability with the efficient computation in case of linear relationships between variables, and neural networks can successfully discover the complex nonlinearity by the multilayered structure with the ability of auto feature extraction. Stochastic Gradient Descent facilitates efficient parameter optimization for large-scale datasets through incremental updates. With this variety of algorithms up our sleeve, we are capable of selectively selecting the best-suited model according to our needs for interpretability, computational performance, and predictive accuracy.

Table 1
Summary of the machine learning models used and their algorithm hyperparameters.

Model	Algorithm Parameters
GB	- Method: xgboost - Number of trees is 100. - Learning rate is 0.099. - Regularization is 0.6 - Limit depth is 6.
AB	- The number of estimators is 50. - Learning rate is 1. - Regression loss function is linear.
RF	- Number of trees in the forest is 10. - Minimum subset size is 5.
SVMs	- SVM Cost is 1. - Regression loss epsilon: 0.1. - Kernel type is radial basis function. - Numerical tolerance: 0.001 - Iteration limit is 500.
DT	- Minimum instances in leaves are 19. - Minimum subset size: 9. - Maximal tree depth is 100. - The stopping point is at 95% of the majority.
KNN	- Number of nearest neighbors is 5. - Metric is Euclidean. - Weights are uniform.
LR	- None.
NN	- Neurons per hidden layer are 500. - Activation is ReLu. - Solver is L-BFGS. - Regularization is 0.04. - Maximum iterations are 600.
SGD	- loss function is hinge. - Regularization is Lasso (L1). - Learning rate is constant. - Initial learning rate is 0.01. - Number of iterations is 1000.

Hyperparameters are the external model behavior-controlling parameters not directly derivable from the data. These important parameters include the learning rates, the depth for the Random Forest, the regularization strength, and the neural network architecture parameters. The performance of machine learning models is strongly dependent on their architecture and its hyperparameters, with countless possible variations that deeply affect the best predictions accuracy. Properly selected hyperparameters prevent underfitting or overfitting and optimize for model

generalizability and computational efficiency, primarily on the complex dataset where there are many features or a nonlinear relationship exists. Regularization methods are mathematical methods for preventing overfitting by adding a penalty term into the loss function, limiting the complexity of the model during the training. Some popular techniques for regularization are L1 (Lasso), L2 (Ridge), Elastic Net, and dropout for neural networks, each working on different dimensions of model complexity. These techniques are very useful when working with high-dimensional data, where the probability of picking up noise instead of the original signal becomes quite high.

Fig. 4 shows a view of the Pythagorean Forest visualization, which is a view of a collection of 10 decision trees from a Random Forest model, each made as a Pythagorean tree shape having unique branching. The visual encoding is based on geometry to represent key model insights: branch width and brightness represent decision path relevance, and colors mean the span of regression predictions. In contrast to classic forest implementations, this visualization technique allows us to see inside the mechanism of how features lead to model behavior. This method significantly improves the model interpretability and facilitates the communication of sophisticated algorithmic behavior to non-technical stakeholders.

3. Results and discussion

3.1. Model results

The regression model evaluation metrics for all of the developed models are shown in Fig. 5. It offers additional views on prediction accuracy and model fit, each with unique mathematical properties and applications. Mean Absolute Error (MAE) computes the average of absolute differences between forecast and real values, giving intuitive interpretation in the same scale, and is resistant to outliers by treating all errors uniformly. However, MAE does not have context for judging good performance without reference. Mean Squared Error (MSE) calculates the average of the squared difference between predictions and observations, exaggerating the effects of large errors by squaring them, being very sensitive to the outliers, and is good when large prediction errors are especially deadly. However, its main disadvantage lies in the difficulties of interpretation because units are squared. RMSE overcomes the descriptiveness problems of MSE through the square root transformation of MSE to revert to the units of the original, providing a balance between MAE robustness and MSE sensibility to outliers and retaining scale-dependent properties

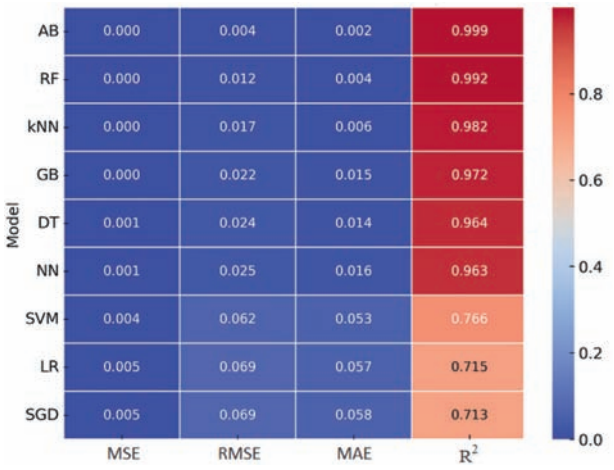


Fig. 5. Developed models performance metrics heatmap.

that make the comparison contextual. Unlike the coefficient of determination (R^2), it contrasts fundamentally by measuring variation explained by the model, thus giving a normalized measure (0–1) that brings inherent context without calling for comparison to other models. R^2 values near to 1 imply high predictive ability, although as a standalone metric, it can promote overfitting because it always grows with any added predictors. This multi-metric strategy allows us to compare absolute and relative performance, to see error distributions, and to conduct informed decisions about model choice based on error tolerance.

The normalized anticipated P4 production rate based on the AB model output that matches the actual P4 production rate is shown in Fig. 6. Although the data points lie just along the 45° straight line in the results, the outputs of the machine learning model are evenly distributed. This means that the constructed AB model led to reliable predictions for P4 production rate values ranging across the entire range of parameters of the data set. Moreover, the accurate forecast of the AB algorithm compared to the actual P4 production rate is also shown in Fig. 7.

Wells of greatest significance in P4 production rate forecasting by the AB model are deployed in the form of SHAP (SHapley Additive exPlanations) in Fig. 8. Table 2 shows the score percentage or f_{ip} factor which is the fraction of injection rate of injector i contributing in production rate of producer p at steady-state condition, based on the SHAP analysis of the AB model. SHAP is a sophisticated method used to explain machine learning models.

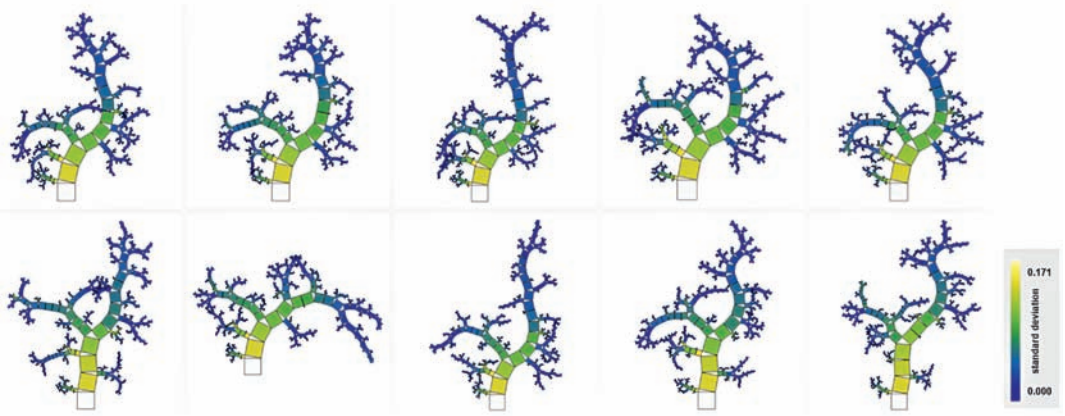


Fig. 4. Pythagorean Forest shows all learned Decision Tree models from the RF model.

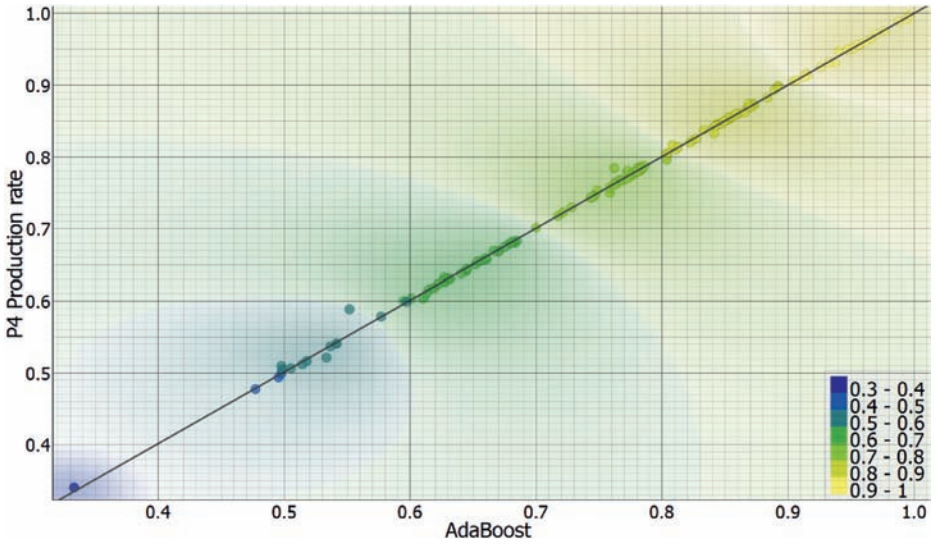


Fig. 6. Predicted vs. actual normalized P4 production rate using AdaBoost model.

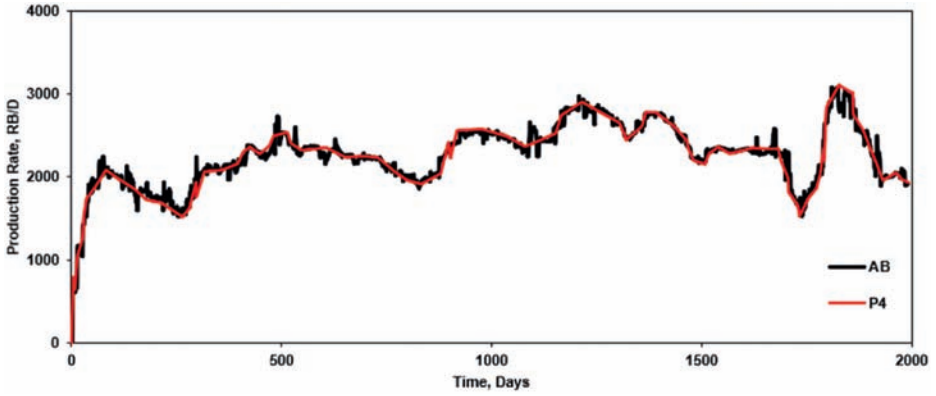


Fig. 7. AdaBoost model match for producer P4.

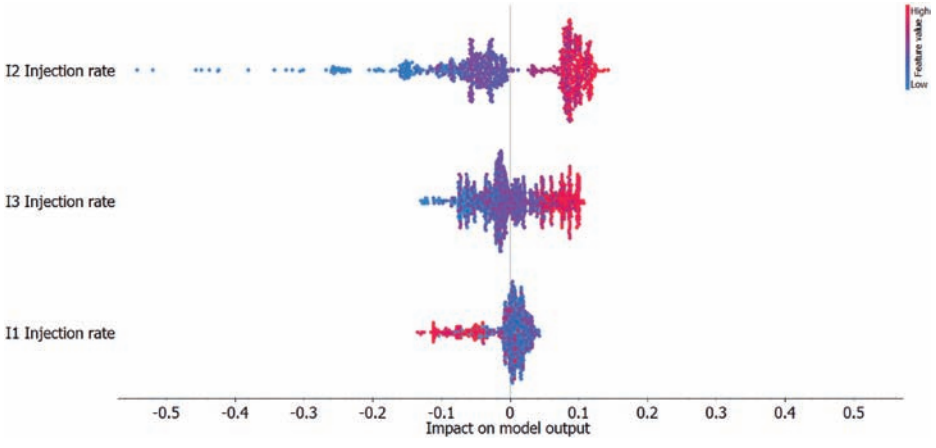


Fig. 8. SHAP plot of the AdaBoost model.

Built on game-theory concepts, SHAP values measure each feature's contribution to particular predictions, delivering both individual and modeled global explanations for behavior. Unlike traditional feature importance methods, SHAP offers significant advantages in interpretability. Unlike the basic feature importance that only ranks variables, SHAP provides the direction of variable effect (positive or negative), deals better with features that are correlated, and gives explanations for particular predictions rather than explanations for the whole of ranking features. The SHAP summary plot displays feature effects as a beeswarm plot where

Table 2
SHAP Analysis scores of the AdaBoost model.

Injection well	Score percentage	F_{ip}
I2 Injection rate	55 %	0.55
I3 Injection rate	28 %	0.28
I1 Injection rate	17 %	0.17

every point renders a Shapley value for a feature and instance, colors represent feature values (red for large and blue for small), and position plot magnitude and direction of influence on the model output. It is model-agnostic and works with any algorithm from simple linear regression to pretty much the most complex deep learning models. This richness of explanation makes SHAP especially useful in high-stakes domains such as waterflooding, where explaining model decisions is numerically important to building trust and ensuring robust implementations. Among the studied wells, I2 has the largest impact on the computed P4 production rate, but others play a bit smaller role.

A contour map generated from SHAP analysis for the AB model is provided in Fig. 9 that represents the magnitude as well as the direction of influence contributed by all injector wells to the production rate of well P4. The clear indication is that the most significant quantity of injected water reaching P4 comes from the southwest direction, implying there was high permeability or channeling between P4 and the southeast area around I2. By drawing a similar contour map for each producer well in the field, a better understanding of interconnectivity between the injector and the producer wells can be gained. This knowledge facilitates the optimization of injection rates for each of the injectors to maximize the overall field production. Such a workflow represents a faster and more accurate method for waterflood optimization, especially for fields with complex reservoirs having large injector-producer networks.

3.2. Model testing and validation

The evaluation of the accuracy of the machine learning model is performed by three methods. The first is the K-fold cross-validation, which is a robust way of model evaluation by splitting the data into K equal-sized subsets, or “folds”. In this method, the model is trained K times in such a way that each time K - 1 number of folds is utilized for training, and then one particular fold is used for validation. The procedure cycles through each fold, making certain that each data point enters the validation set once in all iterations. The last performance metric is then averaged over all K runs. This approach outperforms the simple train-test split in the following aspects: it utilizes the whole dataset for training and validation at the same time, which removes the bias that can be introduced by a random division of the data; the distribution of performance scores is true to the nature of the model variance; and the overfitting is more effectively detected by testing the model's generalizability on many different parts of the dataset. Moreover, K-fold cross-validation offers better performance estimation as well since it reduces the impact of how the data is split, especially useful when working with very limited amounts of data. Even though the computation is more expensive than a simple split, the increased reliability of model evaluation usually makes this worthwhile for uses such as waterflooding optimization, where it is crucial to evaluate model performance. Fig. 10 presents the outcome of the K-fold cross-validation that executed this with 10 folds.

The second validation method is random sampling cross-validation (also called Monte Carlo cross-validation) performs multiple times random splitting of data on the train & test split and averages the performance across the iterations. Unlike K-fold cross-validation, where all the data is divided into K parts of roughly equal size and mutually exclusive such that each point is in the test set exactly once, random sampling provides a flexibility where one can explore different types of data partitions. K-fold has the nearly unbiased estimate of performance, whereas random sampling can lead to more bias if run for the sufficient number of iterations. K-fold ensures that all your data points are exactly used once for testing in a random split, then some points may never be used or they may be used multiple times. Random sampling becomes handy, especially when there is a greater need for control on the training/testing ratio or whenever bothered with extreme outliers disturbing the results, which extremely prohibits K-1 folds in conventional K-fold validation. Both methods effectively prevent overfitting but come from a different statistical perspective on model evaluation. The repeated random sampling method resulted in its outcomes through 10 singular trials, which are listed in Fig. 11. Results include MSE, RMSE, MAE, and R^2 for every model as per the evaluation.

The third validation method we will cover is blind dataset validation, which is a technique that assesses how well a model generalizes to an entirely new dataset not related to the training and validation process. Unlike samples of the same dataset that K-fold cross-validation and random sampling cross-validation will

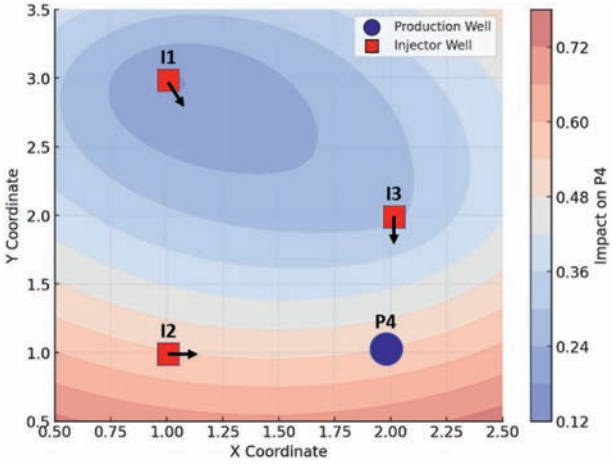


Fig. 9. Contour map showing the impact of injector wells on P4, with the strongest influence from the southwest, indicating high permeability or channeling near I2.

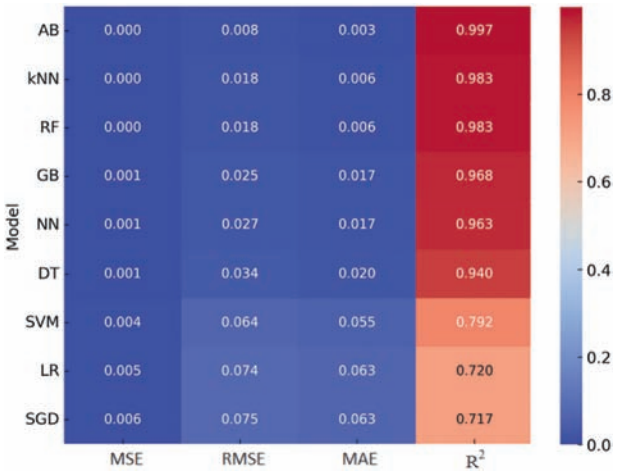


Fig. 10. K-fold cross-validation models performance metrics heatmap.

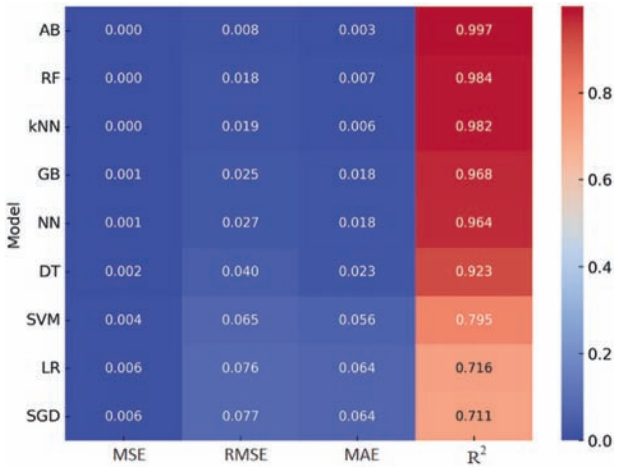


Fig. 11. Random sampling cross-validation models performance metrics heatmap.

reuse, blind validation offers a more authentic measure of performance in real life. It ensures no data leakage and tests the model under different situations. This approach is especially important in high-stakes applications like waterflooding performance evaluation, where real generalization is crucial. While k-fold and random sampling provide valuable insights during tuning the model, it is blind dataset validation that is the final test before actual deployment. We applied the methodology to a completely fresh set of data consisting of the production rate of well P4 and the injection rates of wells I1, I2, and I3. This dataset adds 500 days past the original 2000-day dataset as shown in Figs. 12 and 13. Moreover, Fig. 13 demonstrates the forecast made by the AB algorithm is closely aligned with the P4 future production rate over this period (days 2000–2500 or approximately ±1.4 years). The realized AB model generates excellent accuracy for actual field applications with a mean absolute error of merely 21 RB/D between the anticipated and actual production rates.

4. Conclusion

This study's results show that advanced machine learning models are successfully able to accurately forecast the production of oil wells supported by water injection, measuring the level of communication between injectors and producers, as well as evaluating the efficiency of water used for injection. These machine learning methods can be seen as a near-accurate, faster, less expensive replacement for the current empirical, analytical, capacitance-resistance, and numerical methods. The models were trained against data coming from four wells lasting over 2000 days, achieving outstanding forecasting performance with a minimum RMSE of 0.004 and maximum R^2 of 0.99 obtained by AdaBoost, closely followed by Random Forest (RMSE = 0.012, R^2 = 0.99) and K-Nearest Neighbors (RMSE = 0.017, R^2 = 0.98). Although the Gradient Boosting algorithm performs very slightly less accurately (RMSE = 0.022, R^2 = 0.97), it is yet more accurate than more common simple models like Linear Regression and Decision Tree. The inclusion of feature importance analysis, along with SHAP scores in particular, boosts the model interpretability and quantifies the impact of each injection well on the production well. In addition, the combination of machine learning and SHAP analysis allows building a contour map of the magnitude and direction of the impact of all injector wells on a particular production well in the field. These results suggest that machine learning approaches are suitable for predicting waterflooding performance, helpful in decision-making, and require a smaller number of inputs, costs, and time related to traditional modeling. However, there are some limitations to our approach: the dataset is based on a synthetic reservoir, not a real field; external validation is necessary; the performance of the models is also dependent on the quality of input data and noise; and it will be necessary to test the

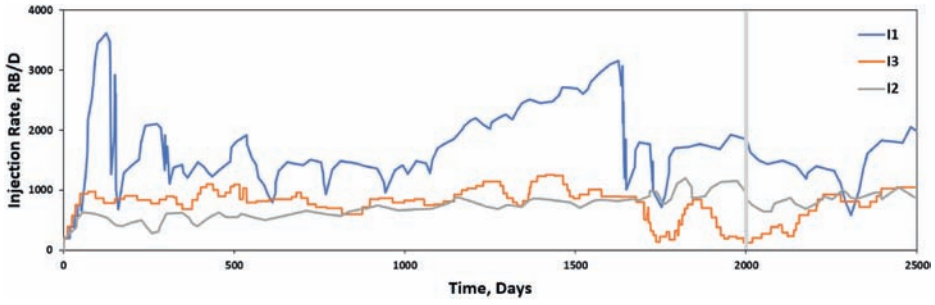


Fig. 12. Injection rates of wells I1, I2, and I3 over 2000 days of the original dataset and an additional 500 days of the unseen dataset.

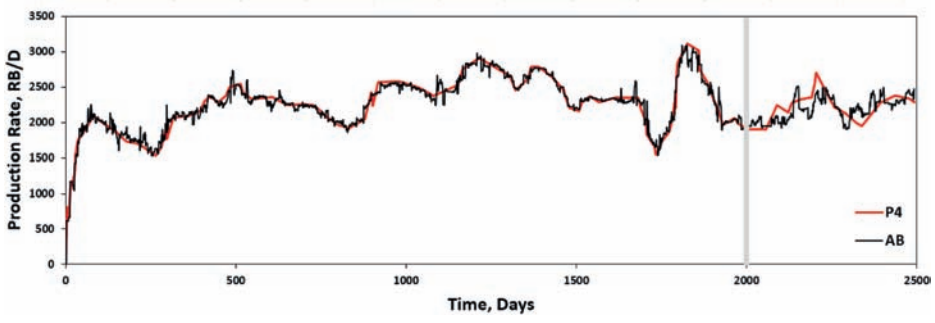


Fig. 13. AdaBoost model match for producer P4 over 2000 days of the original dataset and an additional 500 days of the unseen dataset.

models on larger and more heterogeneous reservoirs. Future studies will be dedicated to the overcoming of these difficulties and the possibility of applying the methodology to the work in the field. This paper has shown the possibility of using state-of-the-art machine learning techniques for real-time waterflood optimization and indicates potential for advanced new data-driven reservoir management techniques.

Funding

We confirm that there are no funding sources to acknowledge for this work.

CRediT authorship contribution statement

Samuel Nashed: Writing – original draft, Visualization, Validation, Methodology. **Oluchi Ejehu:** Formal analysis, Data curation, Conceptualization. **Rouzbeh Moghanloo:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Albertoni, A., Lake, L.W., 2003. Inferring interwell connectivity only from well-rate fluctuations in waterfloods. SPE Reservoir Eval. Eng. 6, 6–16. https://doi.org/10.2118/83381-PA.

Aleev, A., Syrtlanov, V., Kovalyova, Ye, 2017. Operative localization technique of current Mobile oil reserves for mature reservoirs based on material balance. In: SPE Russian Petroleum Technology Conference. SPE, Moscow, Russia. https://doi.org/10.2118/187837-MS.

Alghamdi, A., Hiba, M., Aly, M., Awotunde, A., 2021. A critical review of capacitance-resistance models. In: SPE Russian Petroleum Technology Conference. SPE, Virtual. https://doi.org/10.2118/206555-MS. SPE Russian Petroleum Technology Conference.

Arab, D., Bryant, S.L., Torsæter, O., Kantzas, A., 2022. Water flooding of sandstone oil reservoirs: underlying mechanisms in imbibition vs. drainage displacement. J. Petrol. Sci. Eng. 213, 110379. https://doi.org/10.1016/j.petrol.2022.110379.

Asante, J., Ampomah, W., Carther, M., 2023. Forecasting oil recovery using long short term memory neural machine learning technique. In: SPE Western Regional Meeting. SPE, Anchorage, Alaska, USA. https://doi.org/10.2118/212967-MS.

Batycky, R.P., Thiele, M.R., Razavi, M., 2017. Material balance applied to dynamic reservoir surveillance patterns, in: SPE Western regional meeting. In: SPE Western Regional Meeting. SPE, Bakersfield, California. https://doi.org/10.2118/185713-MS.

Carpenter, C., 2020. Capacitance-resistance model used for integrated detection of water production. J. Petrol. Technol. 72, 51–52. https://doi.org/10.2118/0120-0051-JPT.

Castillo Gamarra, F., Ramos, N., Borsani, I., 2019. Estimation of Mature Water Flooding Performance and Optimization by Using Capacitance Resistive Model and Fractional Flow Model by Layer. IJMO, pp. 303–309. https://doi.org/10.7763/IJMO.2019.V9.727.

Dandan, H., Rui, G., Songhao, H., Yuanbing, W., Zhaowu, Z., 2015. Integrated management and application of horizontal well water flooding technology in a large-scale complicated carbonate oilfield containing high permeability streaks. In: SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition. SPE, Nusa Dua, Bali, Indonesia. https://doi.org/10.2118/176110-MS.

Davidescu, B.-G., Bayerl, M., Puls, C., Clemens, T., 2021. Horizontal versus vertical Wells: assessment of sweep efficiency in a multi-layered reservoir based on consecutive inter-well tracer tests - a comparison between water injection and polymer EOR. In: SPE Europec Featured at 82nd EAGE Conference and Exhibition. SPE, Amsterdam, The Netherlands. https://doi.org/10.2118/205211-MS.

Davudov, D., Malkov, A., Venkatraman, A., 2020. Integration of capacitance resistance model with reservoir simulation. In: SPE Improved Oil Recovery Conference. SPE, Virtual. https://doi.org/10.2118/200332-MS. SPE Improved Oil Recovery Conference.

Ding, B., Guan, B., Liu, W., Chen, B., Sun, J., Li, M., Wu, W., Hua, S., Geng, X., Chen, W., Wang, C., 2021. Mechanism of improving water flooding using the nanofluid permeation flooding system for tight reservoirs in Jilin oilfield. Energy Fuel. 35, 17389–17395. https://doi.org/10.1021/acs.energyfuels.1c02043.

Gabry, M.A., Ramadan, A., Soliman, M.Y., 2025. Innovative method for analyzing inter-well connectivity with continuous wavelet transform. In: Offshore Technology Conference. OTC, Houston, Texas, USA. https://doi.org/10.4043/35951-MS.

Gharieb, A., Adel Gabry, M., Algarhy, A., Elsawy, M., Darraj, N., Adel, S., Taha, M., Hesham, A., 2024a. Revealing insights in evaluating tight carbonate reservoirs: significant discoveries via statistical modeling. An In-Depth analysis using integrated machine learning strategies. In: GOTECH. Presented at the GOTECH. SPE, Dubai, UAE. https://doi.org/10.2118/219199-MS.

Gharieb, A., Elshaafie, A., Gabry, M.A., Algarhy, A., Elsawy, M., Darraj, N., 2024b. Exploring an alternative approach for predicting relative permeability curves from production data: a comparative analysis employing machine and deep learning techniques. In: Offshore Technology Conference. OTC, Houston, Texas, USA. https://doi.org/10.4043/35037-MS.

Gharieb, A., Gabry, M.A., Elsawy, M., Algarhy, A., Ibrahim, A.F., Darraj, N., Sarker, M.R., Adel, S., 2024c. Data analytics and machine learning application for reservoir potential prediction in vuggy carbonate reservoirs using conventional well logging. In: SPE Western Regional Meeting. SPE, Palo Alto, California, USA. https://doi.org/10.2118/218913-MS.

Guo, B.-E., Xiao, N., Martyushev, D., Zhao, Z., 2024. Deep learning-based pore network generation: numerical insights into pore geometry effects on microstructural fluid flow behaviors of unconventional resources. Energy 294, 130990. https://doi.org/10.1016/j.energy.2024.130990.

Hamady, A.E., Gharieb, A., Diaz, A., Bogachev, K., Eydinov, D., 2024. Advanced reservoir permeability distribution prediction: a discrete cosine transform-based history matching workflow. In: ADIPEC. SPE, Abu Dhabi, UAE. https://doi.org/10.2118/222492-MS.

Haryanto, E., Yersaiyn, S., Khattak, I.A., Shah, A.R., 2019. Extending the application of classical conformance plot to assess the water flooding plan. In: Abu Dhabi International Petroleum Exhibition & Conference. SPE, Abu Dhabi, UAE. https://doi.org/10.2118/197395-MS.

He, J., Liu, X., Zhu, X., Jiang, T., He, H., Zhou, L., Liu, Q., Zhu, Y., Liu, L., 2021. Waterflooding characteristics of lithologic reservoir in Ordos basin. Sci. Rep. 11, 2503. https://doi.org/10.1038/s41598-021-82035-4.

Helmy, A.G., Salem, S.K., Elnoby, M., 2024. Sensitivity analysis of hydraulic fracturing parameters for optimum horizontal well spacing in tight oil reservoirs. Arabian J. Sci. Eng. https://doi.org/10.1007/s13369-024-09311-z.

Huang, S., Zhu, Y., Chai, S., Ding, G., Xin, Y., Li, P., 2021. A practical model for water injection evaluation and optimization in typical offshore reservoirs considering formation heterogeneity based on experimental study. In: SPE Middle East Oil & Gas Show and Conference. SPE. https://doi.org/10.2118/204605-MS.

Koray, A.-M., Bui, D., Ampomah, W., Appiah Kubi, E., Klumpenhower, J., 2023. Application of machine learning optimization workflow to improve oil recovery. In: SPE Oklahoma City Oil and Gas Symposium. SPE, Oklahoma City, Oklahoma, USA. https://doi.org/10.2118/213095-MS.

Koroglu, B., Onur, M., Li, Y., Alpak, F.O., 2024. A producer-based capacitance resistance model (CRM) to characterize oil reservoirs with natural water influx: a realistic application. In: SPE Annual Technical Conference and Exhibition. SPE, New Orleans, Louisiana, USA. https://doi.org/10.2118/221062-MS.

Kubota, L., Reinert, D., 2019. Machine learning forecasts oil rate in mature onshore field jointly driven by water and steam injection. In: SPE Annual Technical Conference and Exhibition. SPE, Calgary, Alberta, Canada. https://doi.org/10.2118/196152-MS.

Kumar, M., Swaminathan, K., Rusli, A., Thomas-Hy, A., 2022. Applying data analytics & machine learning methods for recovery factor prediction and uncertainty modelling. In: SPE Asia Pacific Oil & Gas Conference and Exhibition. SPE, Adelaide, Australia. https://doi.org/10.2118/210769-MS.

Lavie, G., Torres, J.A., Kamp, A.M., 2011. An evaluation of criteria for dynamic gridding during reservoir simulation of water flooding and solvent injection. In: SPE Reservoir Simulation Symposium. SPE, The Woodlands, Texas, USA. https://doi.org/10.2118/141796-MS.

Li, K., Horne, R.N., 2003. A decline curve analysis model based on fluid flow mechanisms. In: SPE Western Regional/AAPG Pacific Section Joint Meeting. SPE, Long Beach, California. https://doi.org/10.2118/83470-MS. SPE Western Regional/AAPG Pacific Section Joint Meeting.

Li, Q., Wu, S., Wang, B., Li, X., Li, H., Zhang, J., Meng, L., 2013. A new generation reservoir simulator and its application in a mature water-flooding oilfield. In: SPE Asia Pacific Oil and Gas Conference and Exhibition. SPE, Jakarta, Indonesia. https://doi.org/10.2118/166667-MS.

Li, S., Liu, Y., Xue, L., Yang, L., Yuan, Z., Jian, C., 2021. An investigation on water flooding performance and pattern of porous carbonate reservoirs with bottom water. J. Petrol. Sci. Eng. 200, 108353. https://doi.org/10.1016/j.petrol.2021.108353.

Li, X., Zhang, J., Liu, H., Luo, X., Du, K., Hou, B., Wang, L., 2024. A comprehensive machine learning algorithm based methodology for intelligent water injection optimization - case study of D field in Asia. In: APOGCE 2024. SPE, Perth, Australia. https://doi.org/10.2118/221340-MS.

Liang, X., Weber, D.B., Edgar, T.F., Lake, L.W., Sayarpour, M., Al-Yousef, A., 2007. Optimization of oil production based on A capacitance model of production and injection rates. In: Hydrocarbon Economics and Evaluation Symposium. SPE, Dallas, Texas, U.S.A. https://doi.org/10.2118/107713-MS.

Madhumaya, A., Vyas, A., 2024. Application of machine learning in screening the optimal enhanced oil recovery technique. In: Mediterranean Offshore

Conferenc. SPE, Alexandria, Egypt. https://doi.org/10.2118/223150-MS.
Maucec, M., Saffar, A., Turki, A., Fahmy, I., 2019. Benchmarking of state-of-the-art assisted history matching methods under reservoir uncertainty on a complex water-flooding process. In: International Petroleum Technology Conference. IPTC, Beijing, China. https://doi.org/10.2523/IPTC-19128-MS.
Mohamed, B.A., Aliyev, S., Ibrahim, A.F., Al-Shafloot, T., 2024a. Real-time forecasting of subsurface porosity during drilling using advanced time series models. In: ADIPEC. SPE, Abu Dhabi, UAE. https://doi.org/10.2118/222460-MS.
Mohamed, B.A., Ibrahim, A.F., Abdelgawad, K., Sakthivel, S., 2024b. Methylimidazolium ionic liquids as flowback additives for enhanced hydraulic fracturing. In: ADIPEC. SPE, Abu Dhabi, UAE. https://doi.org/10.2118/222848-MS.
Mokheimer, E.M.A., Hamdy, M., Abubakar, Z., Shakeel, M.R., Habib, M.A., Mahmoud, M., 2019. A comprehensive review of thermal enhanced oil recovery: techniques evaluation. J. Energy Resour. Technol. 141, 030801. https://doi.org/10.1115/1.4041096.
Muggeridge, A., Cockin, A., Webb, K., Frampton, H., Collins, I., Moulds, T., Salino, P., 2014. Recovery rates, enhanced oil recovery and technological limits. Philos. Trans. A Math Phys. Eng. Sci. 372, 20120320. https://doi.org/10.1098/rsta.2012.0320.
Nashed, S., Moghanloo, R., 2025. Replacing gauges with algorithms: predicting bottomhole pressure in hydraulic fracturing using advanced machine learning. Eng 6, 73. https://doi.org/10.3390/eng6040073.
Nashed, S., Lnu, S., Guezei, A., Ejehu, O., Moghanloo, R., 2024. Downhole camera runs validate the capability of machine learning models to accurately predict perforation entry hole diameter. Energies 17, 5558. https://doi.org/10.3390/en17225558.
Niedermaier Custodio, L.F., Carneiro, C. de C., 2024. Global enhanced oil recovery assessment: a comprehensive data-driven analysis using self-organizing maps. Geoenergy Sci. Eng. 237, 212771. https://doi.org/10.1016/j.geoen.2024.212771.
Noirot, J.C., Van Den Hoek, P.J., Zwarts, D., Bjoerndal, H.P., Stewart, G., Drenth, R., Al-Masfry, R., Wassing, B., Saeby, J., Al-Masroori, M., Zarafi, A., 2003. Water injection and water flooding under fracturing conditions. In: Middle East Oil Show. SPE, Bahrain. https://doi.org/10.2118/81462-MS.
Patidar, A.K., Joshi, D., Dristant, U., Choudhury, T., 2022. A review of tracer testing techniques in porous media specially attributed to the oil and gas industry. J. Pet. Explor. Prod. Technol. 12, 3339–3356. https://doi.org/10.1007/s13202-022-01526-w.
Rao, X., 2024. The first application of quantum computing algorithm in streamline-based simulation of water-flooding reservoirs. In: ADIPEC. SPE, Abu Dhabi, UAE. https://doi.org/10.2118/221850-MS. ADIPEC.
Salehi, M.M., Safarzadeh, M.A., Sahraei, E., Nejad, S.A.T., 2014. Comparison of oil removal in surfactant alternating gas with water alternating gas, water flooding and gas flooding in secondary oil recovery process. J. Petrol. Sci. Eng. 120, 86–93. https://doi.org/10.1016/j.petrol.2014.05.017.
Sayarpour, M., Kabir, C.S., Lake, L.W., 2009. Field applications of capacitance-resistance models in waterfloods. SPE Reservoir Eval. Eng. 12, 853–864. https://doi.org/10.2118/114983-PA.
Sayarpour, M., Kabir, C.S., Sepehrnoori, K., Lake, L.W., 2010. Probabilistic history matching with the capacitance-resistance model in waterfloods: a precursor to numerical modeling. In: SPE Improved Oil Recovery Symposium. SPE, Tulsa, Oklahoma, USA. https://doi.org/10.2118/129604-MS.
Shi, X., Zhang, J., Lu, Y., Han, Z., He, Y., 2021. Quantitative evaluation of water flooding in a low resistivity heavy oil reservoir with NMR and conventional logs. In: SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/205928-MS.
Shirazi, M., Mahani, H., Tamsilian, Y., Muggeridge, A., Masihi, M., 2024. Full life cycle review of water-based CEOR methods from pre-injection to post-production. Fuel 356, 129574. https://doi.org/10.1016/j.fuel.2023.129574.
Simha, S., Ujjwal, M., Modi, G., 2021. Practical considerations when using capacitance resistance modelling CRM for waterflood optimization. In: SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition. SPE, Virtual. https://doi.org/10.2118/205650-MS.
Sinha, U., Gautam, S., Dindoruk, B., Abdulwarith, A., 2024. Machine learning-enhanced forecasting for efficient water-flooded reservoir management. In: SPE Improved Oil Recovery Conference. SPE, Tulsa, Oklahoma, USA. https://doi.org/10.2118/218258-MS.
Soromotin, A.V., Martyushev, D.A., Junho Pereira, J.L., 2025. On the application of machine learning algorithms in predicting the permeability of oil reservoirs. Artif Intell Geosci 6, 100126. https://doi.org/10.1016/j.aiig.2025.100126.
Sumnu-Dindoruk, D., Dindoruk, B., 2006. Analytical and numerical solution of nonisothermal buckley-leverett flow including tracers. In: SPE Annual Technical Conference and Exhibition. SPE, San Antonio, Texas, USA. https://doi.org/10.2118/102266-MS.
Temizel, C., Salehian, M., Cinar, M., Gok, I.M., Alklih, M.Y., 2018. A theoretical and practical comparison of capacitance-resistance modeling with application to mature fields. In: SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition. SPE, Dammam, Saudi Arabia. https://doi.org/10.2118/192413-MS.
Temizel, C., Artun, E., Yang, Z., 2019. Improving oil-rate estimate in capacitance/resistance modeling using the Y-Function method for reservoirs under waterflood. SPE Reservoir Eval. Eng. 22, 1161–1171. https://doi.org/10.2118/194497-PA.
Thabet, Samuel A., Elhadidy, A.A., Heikal, M., Taman, A., Yehia, T.A., Elnaggar, H., Mahmoud, O., Helmy, A., 2024a. Next-gen proppant cleanout operations: machine learning for bottom-hole pressure prediction. In: Mediterranean Offshore Conference. SPE, Alexandria, Egypt. https://doi.org/10.2118/223373-MS.
Thabet, S.A., Zidan, H.A., Elhadidy, A.A., Helmy, A.G., Yehia, T.A., Elnaggar, H., Elshiekh, M., 2024b. Application of machine learning and deep learning to predict production rate of sucker rod pump Wells. In: GOTECH. SPE, Dubai, UAE. https://doi.org/10.2118/219231-MS.
Thabet, Samuel Adel, El-Hadydy, A.A., Gabry, M.A., 2024c. Machine learning models to predict pressure at a coiled tubing nozzle's outlet during nitrogen lifting. In: SPE/ICoTA Well Intervention Conference and Exhibition. SPE, The Woodlands, Texas, USA. https://doi.org/10.2118/218294-MS.
Ummah, H.F., Setiati, R., 2022. Determination of polymer injection method to an oil field using screening criteria: literature review. IOP Conf. Ser. Earth Environ. Sci. 1104, 012039. https://doi.org/10.1088/1755-1315/1104/1/012039.
Villamizar, M., Rivas, F., Carvajal, G.A., Md Adnan, F., Kheterpal, A., Knabe, S., Rodriguez, J.A., Al-Jasmi, A., Al-Saad, B., Goel, H.K., 2013. Application of linear and partial correlation techniques to enhance the waterflooding surveillance process. In: SPE Middle East Intelligent Energy Conference and Exhibition. SPE, Manama, Bahrain. https://doi.org/10.2118/167392-MS.
Wang, C., Sun, Z., Sun, Q., Zhang, L., Zhang, X., 2021. Comprehensive evaluation of waterflooding front in low-permeability reservoir. Energy Sci. Eng. 9, 1394–1408. https://doi.org/10.1002/ese3.948.
Wang, J., Qi, X., Liu, Huiqing, Yang, M., Li, X., Liu, Hongguang, Zhang, T., 2022. Mechanisms of remaining oil formation by water flooding and enhanced oil recovery by reversing water injection in fractured-vuggy reservoirs. Petrol. Explor. Dev. 49, 1110–1125. https://doi.org/10.1016/S1876-3804(22)60336-7.
Wang, Y., Han, X., Li, J., Liu, R., Wang, Q., Huang, C., Wang, X., Zhang, L., Lin, R., 2023. Review on oil displacement technologies of enhanced oil recovery: state-Of-The-art and outlook. Energy Fuel. 37, 2539–2568. https://doi.org/10.1021/acs.energyfuels.2c03625.
Xue, L., Liu, P., Zhang, Y., 2023. Status and prospect of improved oil recovery technology of high water cut reservoirs. Water 15, 1342. https://doi.org/10.3390/w15071342.
Yang, Z., Urdaneta, A., 2016. A practical approach to history matching premature water breakthrough in waterflood reservoir simulation: method and case studies in south belridge diatomite waterflood. SPE Reservoir Eval. Eng. 20, 726–737. https://doi.org/10.2118/174006-PA.
Ye, H., Deng, J., Ma, J., Zhang, K., Li, Y., Zhang, H., Zhong, K., 2025. Interwell connectivity analysis method based on injection–production data time and space scale coupling. Processes 13, 373. https://doi.org/10.3390/pr13020373.
Yousef, A.A., Gentil, P.H., Jensen, J.L., Lake, L.W., 2006. A capacitance model to infer interwell connectivity from production and injection rate fluctuations. SPE Reservoir Eval. Eng. 9, 630–646. https://doi.org/10.2118/95322-PA.
Yu, W., Lashgari, H.R., Wu, K., Sepehrnoori, K., 2015. CO_2 injection for enhanced oil recovery in bakken tight oil reservoirs. Fuel 159, 354–363. https://doi.org/10.1016/j.fuel.2015.06.092.
Yuliansyah, A.T., Murachman, B., Purwono, S., 2021. Mathematical model for water flooding and HPAM polymer flooding in enhanced oil recovery. ASEAN J. Chem. Eng. 21, 124. https://doi.org/10.22146/ajche.65531.
Zeinijahromi, A., Borazjani, S., Bedrikovetsky, P., Rodrigues, T., 2014. Low salinity fines-assisted water-flood: analytical modelling and reservoir simulation. In: All Days. Presented at the SPE Asia Pacific Oil & Gas Conference and Exhibition. SPE, Adelaide, Australia. https://doi.org/10.2118/171481-MS.
Zhang, T., Izgec, O., Sayarpour, M., 2018. Heavy oil waterflood application of capacitance resistance models. In: SPE Annual Technical Conference and Exhibition. In: SPE Annual Technical Conference and Exhibition. SPE, Dallas, Texas, USA. https://doi.org/10.2118/191398-MS.