

Artificial Intelligence and Machine Learning Applications in Reservoir Engineering: A Review of Characterization, History Matching, Production Forecasting and Enhanced Oil Recovery

Jan Olsen*

Norwegian University of Science and Technology, Norway

Citation: Olsen J. Artificial Intelligence and Machine Learning Applications in Reservoir Engineering: A Review of Characterization, History Matching, Production Forecasting and Enhanced Oil Recovery. *J Petro Chem Eng* 2026;4(3):423-432.

Received: 14 September, 2026; **Accepted:** 21 September, 2026; **Published:** 23 September, 2026

***Corresponding author:** Jan Olsen, Norwegian University of Science and Technology, Norway

Copyright: © 2026 Olsen J., This is an open-access article published in *J Petro Chem Eng* (JPCE) and distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Artificial intelligence (AI) and machine learning (ML) methods have been applied with increasing frequency across reservoir engineering, driven by the computational cost of physics-based simulation and the accelerating growth of subsurface data. This review consolidates AI/ML applications across four core domains: reservoir characterization, history matching and surrogate reservoir modeling, production forecasting and enhanced and improved oil recovery (EOR/IOR) optimization. Rather than cataloguing individual studies in isolation, it compiles reported performance metrics into a comparative synthesis table to identify recurring methodological patterns and limitations. Across reservoir characterization and EOR screening tasks, ensemble tree-based methods and moderately complex neural architectures performed competitively against more elaborate deep-learning alternatives. In production forecasting, a case study of Wolfcamp A wells in the Midland Basin found that a classical ARIMA model outperformed empirical decline-curve models and deep recurrent architectures (LSTM, GRU) in blind testing on held-out wells, echoed in a companion Bakken Shale study. Physics-informed neural networks, which embed governing flow equations directly into model training, emerged across multiple application areas as a response to the data scarcity constraining purely data-driven approaches in subsurface settings. This review identifies three persistent limitations: inconsistent evaluation metrics across studies, which impede comparison; a documented interpretability gap in increasingly opaque deep-learning architectures, set against the field's reliance on physically justified decision-making; and limited generalizability, with no single model performing reliably across all wells even within one formation. Standardized benchmarking and continued development of hybrid data-physics approaches are identified as priorities for future work.

Keywords: Artificial intelligence; Machine learning; Reservoir engineering; Reservoir characterization; History matching; Production forecasting; Enhanced oil recovery; Decline curve analysis; Physics-informed neural networks; Deep learning

Introduction

Petroleum reservoirs are inherently complex, heterogeneous systems whose behavior is governed by coupled multiphase flow, poorly resolved subsurface geometry and rock and fluid

properties that can only be sampled sparsely through wells, seismic surveys and laboratory core analysis. Predicting how a reservoir will perform under a given development plan has traditionally relied on physics-based numerical simulation. While grounded in first principles, such simulation is computationally

expensive, sensitive to the quality of the underlying geological model and typically requires many iterations of history matching before a model can be trusted for forecasting¹. At the same time, the volume of subsurface data generated across the industry - well logs, three-dimensional seismic volumes, production time series and increasingly real-time sensor streams - has grown to a scale that conventional analytical workflows struggle to fully exploit.

These two pressures, the computational burden of physics-based modelling and the accelerating growth of subsurface data, have driven a rapid expansion of artificial intelligence (AI) and machine learning (ML) applications across the reservoir engineering workflow. Data-driven models have been proposed as faster surrogates for full-physics simulation, as pattern-recognition tools for reservoir characterization from well logs and seismic attributes and as flexible, assumption-light alternatives to empirical decline-curve models for production forecasting². Enhanced and improved oil recovery (EOR/IOR) planning, historically dependent on labor-intensive screening criteria and trial-and-error field pilots, has likewise seen AI applied to recovery-method selection, injection-strategy optimization and recovery-factor prediction³. This growth is not confined to reservoir engineering research alone: recent industry reference works have begun cataloguing hundreds of individual AI case studies spanning the entire upstream value chain, from drilling and well completion through facilities and renewable-energy integration⁴⁻⁷. Within that broader landscape, reservoir engineering - specifically reservoir characterization, history matching and surrogate reservoir modelling, production forecasting and EOR optimization - has arguably received the deepest and most sustained methodological attention, in part because production, pressure and log data are more consistently available and standardized than the drilling or facilities data required for other domains.

Despite this volume of individual studies, the reservoir engineering literature on AI/ML applications remains fragmented along at least three lines. First, relevant studies are scattered across petroleum engineering, geophysics, computer science and general energy journals, with little consolidation of which methods have actually been benchmarked against one another under comparable conditions. Second, reported performance metrics - coefficients of determination, mean squared errors and classification accuracies - are rarely presented in a form that supports cross-study comparison, since individual studies differ in dataset size, formation type and evaluation protocol. Third, the field's growing reliance on inherently opaque deep-learning architectures has outpaced its engagement with interpretability, even though reservoir engineering decisions - well placement, EOR method selection, capital allocation - carry financial and operational stakes that typically demand an explainable basis⁸.

This review addresses these gaps by narrowing its scope specifically to AI/ML applications in reservoir engineering, deliberately excluding drilling, well completion, facilities and renewable-energy applications, which are treated at length elsewhere⁴⁻⁷, in order to allow deeper treatment of four core application areas: reservoir characterization, history matching and surrogate modelling, production forecasting and EOR/ IOR optimization. Rather than simply cataloguing individual studies, this review consolidates reported performance metrics across studies into a comparative synthesis table organized by

method, application area and dataset characteristics and uses that consolidated view to ground a critical discussion of the field's current limitations, including data scarcity, limited cross-reservoir generalizability and the interpretability gap associated with black-box models. The review closes by outlining where physics-informed and hybrid data-physics approaches are beginning to address these limitations and where methodological standardization is still needed before AI-based reservoir engineering tools can be adopted with the same confidence afforded to conventional physics-based methods.

AI/ML Methods: A Primer

Three broad families of methods recur across the reservoir engineering applications reviewed here and a brief orientation to each is useful before turning to specific application areas.

Classical machine learning. Regression- and tree-based methods remain the workhorses of reservoir engineering ML, particularly for tabular, structured data such as well-log measurements, core-derived petrophysical properties and production summary statistics. Random forests, an ensemble of decision trees trained on bootstrapped data subsets, offer robustness to overfitting and have been widely adopted for classification and regression tasks in facies identification and permeability prediction⁹. Gradient-boosting methods such as XGBoost build trees sequentially, with each new tree correcting the residual error of its predecessors and have become a common benchmark for reservoir property prediction tasks because of their strong performance on moderate-sized tabular datasets¹¹. Support vector machines, which construct a maximum-margin hyperplane to separate classes (or, in regression form, to fit within a margin of tolerance), remain popular for facies and lithology classification problems, especially when training data are limited relative to the number of input features¹¹. These classical methods generally require less data and offer more interpretable decision boundaries than deep learning alternatives, which explains their continued use even as deep learning has become more prominent elsewhere in the field.

Deep learning. Convolutional neural networks (CNNs) originally developed for image recognition, are well suited to spatially structured subsurface data such as seismic volumes and petrographic thin-section images, where local spatial patterns (fault geometries, facies boundaries, pore structures) carry the relevant signal. For sequential data such as production time series, recurrent architectures are the standard choice. Long short-term memory (LSTM) networks address the vanishing-gradient problem that limits standard recurrent networks' ability to learn long-range temporal dependencies, using gated memory cells to retain or discard information across time steps¹². Gated recurrent units (GRUs) simplify the LSTM gating structure while retaining much of its capacity to model temporal dependencies and have been applied directly to production decline-curve analysis in gas reservoirs¹³. Both architectures appear throughout the production-forecasting literature reviewed in a later section, typically benchmarked against classical time-series models such as ARIMA and against empirical decline-curve models.

Physics-informed and hybrid approaches. A more recent development addresses a limitation common to both classical ML and deep learning: purely data-driven models can produce predictions that are statistically plausible but physically inconsistent, particularly when extrapolating beyond the range

of training data. Physics-informed neural networks (PINNs) address this by embedding the governing partial differential equations of the system, such as the diffusivity equation or the Buckley-Leverett equation for two-phase flow, directly into the network's loss function alongside the data-fitting term, so that the trained model is penalized for violating known physical constraints¹⁴. In reservoir engineering specifically, this approach has been extended to handle the sparse, unevenly distributed production data typical of real fields, for example through domain-decomposition schemes that separate near-well regions (where pressure gradients are steep and data are relatively dense) from the surrounding reservoir volume (where data are sparse but gradients are gentler)¹⁵. Because subsurface data scarcity is one of the recurring limitations discussed later in this review, physics-informed and hybrid data-physics approaches recur as a proposed remedy across several of the application areas that follow.

Reservoir Characterization

Reservoir characterization, the task of inferring rock and fluid properties away from direct measurement points, has attracted more machine-learning research than perhaps any other reservoir engineering subdomain, in part because well-log and seismic datasets are comparatively abundant and already digitized across the industry.

Facies classification

Lithofacies classification is a natural fit for supervised classification methods, since well-log or core-derived facies labels provide ground truth against which classifier performance can be directly measured. Narayan et al. compared six supervised classifiers - random forest, artificial neural network, adaptive boosting, extreme gradient boosting, support vector machine and multilayer perceptron - for classifying shale, sand and carbonate litho-facies from 3D seismic attributes in the Penobscot field, Scotian Basin, using well-log-derived facies as training labels. The multilayer perceptron achieved the highest overall classification accuracy (87.7%), followed closely by random forest (86.1%) and artificial neural network (85.3%), with extreme gradient boosting (84.9%), support vector machine (81.4%) and adaptive boosting (79.2%) trailing behind; all six methods performed markedly better on the shale facies (86-94% accuracy) than on sand (65-79%) or carbonate (60-78%) facies, a pattern the authors attributed to the relative thickness and sample abundance of each facies type in the training data¹⁶. More recent work has moved toward deeper architectures: Nguyen, Nguyen and Mai proposed an attention-based residual neural network for facies classification, reporting an area under the receiver operating characteristic curve of 0.883 and found that the attention mechanism's ability to focus on the most informative log intervals improved performance over both single-stream convolutional networks and conventional machine-learning baselines¹⁷. At a more general methodological level, Xie, et al. compared several machine-learning methods for formation lithology identification and emphasized that reported accuracy differences between methods were frequently smaller than the variation introduced by hyperparameter tuning choices, a caution echoed by Li, et al., who applied a semi-supervised Laplacian support vector machine to lithology identification specifically to reduce dependence on the large volumes of labelled data that fully supervised methods require^{18,19}. Antariksa, Muammar

and Lee combined machine-learning classification with rock-physics analysis for lithofacies prediction in the Tarakan Basin, Indonesia, arguing that incorporating physically motivated rock-physics features alongside raw log data improved classification robustness relative to log-only feature sets²⁰.

Petrophysical property prediction

Beyond discrete facies labels, a parallel body of work has applied regression-based machine learning to continuous petrophysical properties. Yang, Dindoruk and Lu compared several machine-learning algorithms against classical empirical correlations for bubble-point pressure prediction and reported that the machine-learning models achieved lower prediction error across a wider range of fluid compositions than any single classical correlation, though at the cost of reduced extrapolation reliability outside the training data's compositional range²¹. Anemangely et al. applied machine learning to predict shear-wave velocity from petrophysical logs in wells lacking direct shear-sonic measurements, a common data gap in older well files and found that gradient-boosted and neural-network models outperformed the classical empirical correlations typically used for this purpose²². Water saturation, a property central to both original-oil-in-place estimation and reservoir characterization more broadly, has similarly been targeted: Khan, Tariq and Abdulraheem developed a machine-learning-derived correlation for water saturation specifically in complex, mixed-lithology reservoirs where classical Archie-type equations perform poorly²³. At the pore scale, Alqahtani et al. applied machine learning directly to two-dimensional X-ray images of porous media to predict properties such as porosity and permeability, illustrating how imaging-based characterization is beginning to extend machine-learning applications below the log scale to the pore scale²⁴.

Seismic interpretation

Seismic data, while lower in vertical resolution than well logs, offers the only source of areal coverage away from well control and deep learning has increasingly been applied to extract structural and stratigraphic information from seismic volumes directly. Cunha, et al. applied a convolutional neural network pre-trained on synthetic seismic data via transfer learning to fault detection in real seismic volumes, addressing the practical difficulty that manually labelled real fault data are scarce relative to the volume of seismic data available for training²⁵. This transfer-learning strategy - pre-training on synthetic or simulated data before fine-tuning on the smaller volume of labelled field data - recurs as a proposed solution to data scarcity across several of the reservoir characterization studies discussed above and reappears in the discussion of current limitations later in this review.

History Matching and Surrogates

Deep-learning surrogate models

History matching, the process of adjusting an uncertain reservoir model's parameter until its simulated output matches observed production and pressure data, is fundamentally an inverse problem that in its classical form requires running the full numerical simulator many hundreds or thousands of times across an ensemble of candidate models. Because each simulation run can take from minutes to hours depending on model size, this computational burden is the central motivation

for developing faster surrogate models, a point already introduced in this review's discussion of the field's underlying pressures¹. Shahkarami and Mohaghegh describe this class of methods under the umbrella term "smart proxies" - data-driven models trained to reproduce the input-output behaviour of a full-physics simulator across the range of geological and operational scenarios of interest - and report their application across reservoir simulation, well-placement optimization and production-decline problems²⁶. Tang, Liu and Durlafsky developed a deep-learning-based surrogate specifically for data assimilation in dynamic subsurface flow problems, using a convolutional architecture to map from geological parameter fields directly to simulated pressure and saturation outputs and later extended this approach to full three-dimensional history matching using a CNN-PCA geological parameterization that reduces the dimensionality of the geological uncertainty space before surrogate training^{27,28}. Alakeely and Horne applied convolutional and recurrent neural networks directly to simulate reservoir behavior from historical production and injection data, reporting that the trained networks could reproduce simulator-level dynamic behavior at a fraction of the computational cost once trained, though the upfront training-data-generation cost (running the full simulator to generate training examples) remains a limiting factor for very large or geologically complex models²⁹.

Optimization-integrated history matching

Beyond standalone surrogate models, several studies have integrated machine learning directly into the history-matching optimization loop itself. Cavalcante et al. proposed a stochastic learning-from-data approach that treats history matching as a pattern-recognition problem, learning the relationship between reservoir parameters and observed data mismatch directly from an ensemble of prior simulation runs rather than iterating a single model toward convergence³⁰. Chai, et al. developed an integrated closed-loop workflow that combines assisted history matching with field development optimization using machine-learning surrogates at both stages, arguing that decoupling these two steps, as is common practice, discards information about how a given history-matched model performs under the specific development scenarios an operator ultimately cares about³¹. At the well-design level, Kim, Yang and Choe used convolutional-neural-network-based proxy models to optimize the locations and types of multiple wells under geological uncertainty, replacing repeated full-simulation evaluations within the optimization loop with fast proxy evaluations, while Wang, et al. proposed a distributed quasi-Newton optimization workflow for well-location optimization designed specifically to reduce the number of expensive simulation calls required to reach a bi-objective (economic and technical) optimum^{32,33}. A related but distinct problem, automatic interpretation of well-test pressure-transient data, has also been addressed with convolutional networks: Li, et al. applied a CNN directly to well-test pressure derivative curves for a radial composite reservoir, framing flow-regime identification as an image-classification problem rather than the manual type-curve matching traditionally used by well-test analysts³⁴.

Physics-informed extensions

As introduced earlier in this review, physics-informed neural networks address a specific limitation of the purely data-driven surrogate models described above: their outputs are

not guaranteed to satisfy the underlying flow physics, which becomes especially problematic when a surrogate is asked to extrapolate to parameter combinations or well configurations outside its training data. Han et al.'s domain-decomposition PINN, introduced earlier as an example of the physics-informed approach, was developed specifically to address this limitation for large-scale reservoir simulation, using the sparse production and pressure data actually available from real wells rather than requiring dense synthetic training data generated by running a full simulator repeatedly.¹⁵ Because generating adequate training data is itself often the single most expensive part of building a conventional data-driven surrogate, this ability to train directly on sparse field data, rather than on simulator output, represents one of the more practically significant advantages physics-informed approaches offer over the purely data-driven surrogate models reviewed above and is a theme this review returns to in its discussion of future directions.

Production Forecasting

Empirical decline-curve foundations

Accurate forecasting of well production underlies nearly every reservoir management decision, from capital allocation to enhanced-recovery timing and nowhere is this more consequential than in unconventional reservoirs, where primary recovery factors typically remain below 10% and production declines steeply within the first months on stream. The dominant traditional approach, decline-curve analysis (DCA), fits a small number of parameters to historical rate-time data and extrapolates the fitted curve forward. The original Arps formulation, comprising exponential, hyperbolic and harmonic decline forms, remains the most widely used DCA framework nearly eight decades after its introduction³⁵. However, Arps' hyperbolic model was derived for boundary-dominated flow in conventional reservoirs and its extrapolation to unconventional wells, which spend much of their productive life in transient or linear flow regimes governed by complex, multi-stage hydraulic fracture networks, has repeatedly been shown to produce unrealistic decline exponents and to overestimate estimated ultimate recovery³⁶. This limitation motivated a family of modified empirical models, including Duong's rate-decline model for fracture-dominated flow³⁶ and the logistic growth model, which imposes a finite ultimate-recovery ceiling on the fitted curve and has performed competitively against other empirical forms in several unconventional benchmarking studies³⁷.

Machine-learning alternatives

Where empirical DCA models fit a small number of physically motivated parameters, machine-learning approaches instead learn a flexible mapping between historical production (and, in some formulations, additional geological or operational variables) and future rates, without committing to a fixed functional form. Early applications used classical artificial neural networks; Al-Fattah and Startzman, for example, applied an ANN to predict U.S. natural gas production trends using historical production data as the sole input³⁸. More recent work has shifted toward architectures purpose-built for sequential data. Long short-term memory (LSTM) networks have been applied directly to petroleum production time series, exploiting their capacity to retain long-range dependencies in noisy, non-stationary rate data³⁹, while hybrid architectures that combine fiber-optic distributed sensing data with multilayer

perceptrons have been used to model gas production specifically in the Marcellus Shale⁴⁰. Gated recurrent units have likewise been applied to decline-curve analysis in gas reservoirs, offering comparable sequence-modeling capacity to LSTM at lower computational cost¹³.

Comparative and benchmarking studies

Because both empirical and ML approaches are now well established, a growing body of work has directly benchmarked them against one another rather than proposing either in isolation. Sun, Ma and Kazi compared conventional DCA against recursive neural networks for multi-well production forecasting and found the neural-network approach more robust to noisy, multi-well datasets, though at the cost of interpretability⁴¹. Panja, et al. similarly found that AI-based forecasts outperformed conventional DCA for shale wells, particularly over longer forecast horizons where DCA's fixed functional form becomes a binding constraint⁴². At a broader methodological level, Ning, Kazemi and Tahmasebi compared ARIMA, LSTM and Prophet (a decomposable time-series model originally developed for business forecasting) across oil production datasets and found that ARIMA and LSTM generally outperformed Prophet, although Prophet captured seasonal production fluctuations more effectively where they were present⁴³. Tadjer, Hong and Bratvold likewise combined classical decline-curve analysis with machine-learning-based short-term forecasting and reported that the hybrid approach mitigated the tendency of any single decline model to over- or under-estimate production during the early, most volatile phase of a well's life⁴⁴.

Case Study: Wolfcamp A, Midland Basin

A recent study by Laalam, et al. illustrates both the promise and the nuance of this comparative literature in unconventional practice⁴⁵. Using production data from eleven wells in the Wolfcamp A interval of the Midland Basin, the authors benchmarked eleven empirical decline-curve models, including the Arps exponential, hyperbolic and harmonic forms, the Weibull model, the logistic growth model and several extended hybrid decline forms, against ARIMA, LSTM and GRU time-series models trained on eight of the eleven wells, with the remaining three reserved for blind testing. Among the empirical models, the logistic growth model, the generalized Weng model and the Wang model consistently ranked highest across all eight training wells, while models such as the Weibull and Arps harmonic forms showed markedly inconsistent rankings from well to well, underscoring that no single empirical form generalizes reliably even within a single formation. On the time-series side, a per-well hyperparameter-optimized ARIMA model achieved R² scores ranging from 72.0% to 90.5% across the eight training wells; when the individual well forecasts were aggregated into a single combined model, the ARIMA ensemble achieved an R² of 93.5% against the averaged production data, compared with 65.9% for the best-performing empirical model (the logistic growth model) over the same aggregated comparison. In blind testing on the three held-out wells, ARIMA again led, with an average R² of 80.0%, ahead of GRU (75.5%) and LSTM (62.6%); the authors attributed LSTM's weaker blind-test performance to gaps in fit at the start and end of the production history, suggesting a need for further hyperparameter tuning rather than a fundamental limitation of recurrent architectures. A companion study by the same lead author applying the same

comparative framework to Bakken Shale wells in the Williston Basin reported a similar pattern: no single model dominated across all wells, but a properly tuned ARIMA model performed competitively for the majority of them⁴⁶.

Toward hybrid data-physics models

The most recent work in this area has begun to blend the empirical and machine-learning traditions more directly rather than treating them as competing alternatives. Matoorian, et al. introduced a hybrid data-physics architecture that embeds physics-based decline equations within a deep-learning framework, reporting improved accuracy over purely data-driven models for both short- and long-term forecasts in the Duvernay play while requiring less training data⁴⁷. Similarly, Chen, et al. combined convolutional and gated recurrent layers with Bayesian hyperparameter optimization to forecast production in an enhanced-oil-recovery setting, extracting both spatial and temporal structure from multivariate operational data⁴⁸, while Du, et al. incorporated an autoregressive component alongside a modified GRU architecture specifically to account for external economic drivers of production decisions⁴⁹. More broadly, several of the deep-learning and hybrid forecasting approaches reviewed above are increasingly framed within a wider Industry 4.0 and smart-field context, in which production-forecasting models function as one component of integrated workflows spanning oilfield development planning, well optimization and hydrocarbon production forecasting together, rather than as standalone forecasting tools deployed in isolation⁵⁸. Taken together, this recent work echoes the physics-informed direction introduced earlier in this review: purely data-driven sequence models have repeatedly demonstrated strong average performance, but the field's more careful benchmarking studies, particularly those using blind testing on genuinely unseen wells, continue to find that no single model, empirical or data-driven, generalizes reliably across all wells within a formation - itself one of the central findings this review returns to in its discussion of current limitations.

EOR and IOR Optimization

EOR method screening

Selecting an appropriate enhanced-recovery method for a given candidate reservoir has traditionally relied on published screening criteria and analogous-field comparisons, a process that is time-consuming and sensitive to the specific set of analog fields a given engineer happens to be aware of. Machine learning has been applied directly to this screening decision as a multi-class classification problem: Su et al. compiled a worldwide dataset of enhanced-recovery projects and trained random forest, artificial neural network, naïve Bayes, support vector machine and decision-tree classifiers to recommend an EOR category (thermal, gas injection or chemical flooding) and specific technique given reservoir and fluid properties, with the random forest classifier achieving the highest prediction accuracy of the methods tested: 100% for thermal-flooding techniques, 96.5% for gas-injection techniques and 88.5% for chemical-flooding techniques⁵⁰. The authors attributed the comparatively lower chemical-flooding accuracy to greater within-category heterogeneity among chemical EOR techniques (polymer, surfactant and alkaline flooding each behave differently) relative to the thermal category, where techniques such as steam flooding are more mechanistically similar to

one another. A recent, broader review by Jing, et al. surveyed machine-learning applications across four EOR categories - chemical flooding, gas injection, thermal recovery and emerging microbial flooding - and identified four recurring cross-cutting challenges: the need for closer interdisciplinary collaboration between reservoir engineers and data scientists, the difficulty of multiscale modelling (reconciling pore-scale mechanisms with field-scale predictions), model interpretability and inconsistent data quality across studies, all four of which recur, in reservoir-engineering-specific form, in this review’s own discussion of current limitations below³.

Injection and production optimization

Beyond upfront method selection, machine learning has also been applied to optimize the ongoing operational parameters of an EOR project once a method has been selected. Sun and Ertekin developed artificial-neural-network-based proxies specifically for screening and optimizing polymer-flooding project designs, allowing rapid evaluation of injection schedules that would otherwise each require a separate full-physics simulation run⁵¹. For waterflooding, a more mature but still operationally complex recovery method, Zhong, et al. built a machine-learning-based proxy model to predict field-level production rates directly from injection and operational parameters, while Negash and Yaw applied an artificial neural network specifically to forecast production under water injection in a hydrocarbon reservoir, reporting improved accuracy relative to conventional material-balance-based forecasting approaches for the field studied^{52,53}. For CO2-based EOR specifically, which carries the additional complexity of miscibility behaviour and, increasingly, carbon-storage accounting alongside recovery optimization, Karacan applied a fuzzy-logic approach to estimate recovery factors for

miscible CO2-EOR projects across the United States, while You et al. developed a machine-learning-assisted workflow to optimize water-alternating-CO2 injection field operations, reporting that the ML-assisted workflow identified injection strategies that outperformed the operator’s baseline scheme in simulated net-present-value terms^{54,55}. Addressing the class-imbalance problem that arises naturally in EOR datasets, where some techniques (such as waterflooding) are represented by far more historical projects than others (such as microbial EOR), Pirizadeh, et al. proposed an ensemble machine-learning model specifically designed to improve classification performance on underrepresented EOR categories, a data-quality issue that echoes the imbalance concern Su et al. also noted in their own screening dataset⁵⁶. At the level of full-field operational optimization under uncertainty, Moeini and Babaei combined support vector machines with a chaotic-integrated particle swarm optimization algorithm to optimize reservoir operation without requiring advance knowledge of future operating conditions, illustrating how classical optimization metaheuristics continue to be paired with machine-learning surrogates even as deep-learning-based alternatives have become more common elsewhere in the literature⁵⁷.

Comparative Synthesis

(Table 1) consolidates the reported performance metrics from the studies above that provided sufficiently specific, directly comparable numbers to be tabulated. It is not exhaustive of every study reviewed above; rather, it selects the entries where a study directly compared two or more methods on the same dataset using a clearly stated metric, since these are the entries from which a genuine methodological comparison can be drawn.

Table 1: Comparative performance of AI/ML methods across reservoir engineering application areas.

Application Area	Study	Methods Compared	Best-Performing Method	Reported Metric	Dataset
Reservoir characterization	Narayan, et al. ¹⁶	RF, ANN, ADB, XGB, SVM, MLP	MLP	87.7% overall accuracy	3D seismic + logs, Penobscot Field
Reservoir characterization	Nguyen, et al. ¹⁷	Attention-ResNet vs. conventional ML/DL	Attention-ResNet	AUROC 0.883	8-well dataset, mixed geology
Production forecasting (aggregated)	Laalam, et al. ⁴⁵	11 empirical DCA models, ARIMA, LSTM, GRU	ARIMA (aggregated)	R ² 93.5% vs. 65.9% (best empirical)	8 wells, Wolfcamp A, Midland Basin
Production forecasting (blind test)	Laalam, et al. ⁴⁵	ARIMA, LSTM, GRU	ARIMA	R ² 80.0% vs. GRU 75.5%, LSTM 62.6%	3 held-out wells, Wolfcamp A
EOR screening	Su, et al. ⁵⁰	RF, ANN, naïve Bayes, SVM, decision tree	RF	Accuracy 100%/96.5%/88.5% (thermal/gas/chemical)	Worldwide EOR project database

Three observations follow from this table, each of which anticipates the discussion of limitations below. First, ensemble tree-based methods and moderately complex neural architectures (random forest, multilayer perceptron) consistently rank at or near the top across both reservoir characterization and EOR screening tasks, rather than being displaced by more elaborate deep-learning alternatives. Second and somewhat contrary to the general trend toward deep learning elsewhere in the literature, the classical ARIMA time-series model outperformed both empirical decline-curve models and deep recurrent architectures (LSTM, GRU) in the one production-forecasting study reviewed here that included genuine blind testing on held-out wells. Third, the metrics themselves are heterogeneous - accuracy, AUROC and R² are not directly convertible into one another - which

means a table like this one can juxtapose results but cannot fully resolve the underlying comparability problem, since no two studies in this literature use the same wells, formation or evaluation protocol.

Critical Gaps and Challenges

Three limitations recur across the application areas reviewed above and each has been noted, in one form or another, by multiple independent groups of authors rather than being a single reviewer’s idiosyncratic concern.

Inconsistent evaluation and reporting

As the comparative synthesis above illustrates, studies across reservoir characterization, history matching, production

forecasting and EOR optimization report performance using different metrics, on different datasets, with different train-test splits and, in some cases, without any held-out blind test at all. Where blind testing on genuinely unseen wells has been performed, as in the Wolfcamp A study reviewed above, results have sometimes diverged from the same models' training-set performance, underscoring that in-sample accuracy figures alone can overstate a model's practical reliability⁴⁵. No standard benchmark dataset or evaluation protocol has emerged across reservoir engineering ML research comparable to, for example, the shared benchmark datasets that have driven comparability and progress tracking in computer vision or natural language processing; the closest analogs are formation-specific public datasets such as the Penobscot field seismic-and-log dataset used in several of the studies reviewed above, but these remain the exception rather than the norm.

Data scarcity and quality

Labelled subsurface data - whether facies interpretations, history-matched geological realizations or field-verified EOR outcomes - are expensive to generate and, in the case of proprietary field data, frequently unavailable for public benchmarking at all. Several of the studies reviewed above address this directly: transfer learning from synthetic seismic data was used specifically to compensate for the scarcity of labelled real fault data²⁵, semi-supervised learning was applied to lithology classification specifically to reduce dependence on fully labelled training sets¹⁹ and physics-informed neural networks were extended specifically to operate on the sparse production data realistically available from field wells rather than requiring dense simulator-generated training data¹⁵. That three independent lines of work, spanning seismic interpretation, log-based classification and physics-informed simulation, converged on data scarcity as a central obstacle suggests this is a structural feature of the domain rather than an artifact of any single dataset.

The interpretability gap

Reservoir engineering decisions - well placement, EOR method selection and capital-intensive field development commitments - carry financial and operational consequences that typically require an engineer to justify a recommendation on physical or economic grounds, not merely to report a model's aggregate accuracy. Yet the field's more recent methodological direction has favoured increasingly opaque deep-learning architectures over the more interpretable tree-based and linear methods that dominated earlier work. Explainable-AI techniques developed outside petroleum engineering, such as post hoc feature-attribution methods, offer a partial remedy, but their adoption within the reservoir engineering literature reviewed here remains limited relative to their use in other high-stakes domains such as healthcare and finance⁸. Physics-informed approaches offer an alternative, structural route to interpretability, since a model constrained to satisfy known governing equations is, by construction, harder to render physically nonsensical than an unconstrained black-box network, even if the internal weights of the network itself remain no more interpretable than before.

Limited cross-reservoir generalizability

Finally and perhaps most directly relevant to practicing reservoir engineers deciding whether to trust a published method for their own field, the comparative studies reviewed in the

production-forecasting section above found that no single model, empirical or machine-learning-based, generalized reliably across all wells even within a single formation^{45,46}. If model rankings are unstable across wells within one formation, extrapolating a method's reported performance to a different formation, basin or geological setting entirely should be treated with particular caution - a point this review's own comparative table above is not able to fully resolve, since the studies compiled there differ in formation, dataset and evaluation protocol in ways that a single table cannot normalize away.

Future Directions

Building on the limitations identified above, three directions appear best positioned to advance AI/ML applications in reservoir engineering over the coming years.

First, hybrid data-physics architectures, already introduced in this review's discussion of physics-informed neural networks and revisited in the history-matching and production-forecasting sections above, are likely to see continued development specifically because they address the data-scarcity limitation directly rather than working around it. Matorian, et al.'s hybrid architecture for unconventional production forecasting and Han et al.'s domain-decomposition PINN for large-scale reservoir simulation both illustrate a common pattern: embedding known governing equations reduces the volume of labelled training data a model needs to reach a given level of accuracy, which is precisely the resource most difficult to obtain in subsurface applications^{47,15}. Continued work extending this hybrid approach to reservoir characterization and EOR optimization, where it remains comparatively less developed than in production forecasting and history matching, appears a natural next step.

Second, the absence of standardized benchmark datasets and evaluation protocols, identified above as a central limitation, is addressable in principle, even if doing so would require coordinated effort across the research community (through a professional society such as the Society of Petroleum Engineers, for example) rather than any single research group's initiative. The Penobscot field dataset's repeated use across multiple independent characterization studies reviewed above suggests that shared, public datasets, when they exist, do become de facto benchmarks; deliberately curating and releasing additional shared datasets specifically for production forecasting and EOR screening comparisons - the two application areas where dataset heterogeneity across studies was most apparent in this review's comparative table - would allow future comparative studies to isolate genuine methodological differences from dataset-specific artifacts.

Third, the increasing availability of real-time subsurface and surface sensor data, noted at the outset of this review as one of the two structural pressures driving AI/ML adoption in the first place, points toward a shift from static, retrospectively trained models toward models that continuously update as new field data arrive. Several of the surrogate-modelling and history-matching studies reviewed above already move in this direction implicitly, since a surrogate model's value is highest precisely when it can be queried repeatedly as new data justify updating a reservoir model, rather than being trained once and used indefinitely thereafter. Realizing this shift fully will likely require closer integration between the AI/ML methods reviewed in this paper and the industrial data infrastructure - historian

systems, IoT sensor networks and cloud computing platforms - that determine whether continuously updating models are operationally feasible at field scale, a topic that falls outside reservoir engineering narrowly defined but will likely determine how quickly the methods reviewed here move from published case studies into routine field practice.

Conclusion

This review has surveyed artificial intelligence and machine learning applications across four core reservoir engineering domains - reservoir characterization, history matching and surrogate reservoir modelling, production forecasting and enhanced and improved oil recovery optimization - drawing on both peer-reviewed journal literature and an extensive set of documented industry case studies. Several patterns recur across all four domains. Ensemble tree-based methods and moderately complex neural architectures consistently perform competitively against more elaborate deep-learning alternatives on classification tasks, while for production forecasting specifically, the comparative literature reviewed here, including a detailed case study of Wolfcamp A wells in the Midland Basin, found that classical ARIMA time-series models have repeatedly matched or outperformed deep recurrent architectures in properly blind-tested comparisons. Physics-informed and hybrid data-physics approaches have emerged across multiple application areas as a direct response to the data scarcity that constrains purely data-driven models in subsurface settings and appear positioned to see continued development.

At the same time, this review's attempt to consolidate reported performance metrics into a single comparative table also surfaced the limitation that motivated that attempt in the first place: studies across this literature use inconsistent metrics, datasets and evaluation protocols, which makes direct cross-study comparison difficult even when, as in several cases reviewed here, comparable methods have been applied to superficially similar problems. Compounding this, the handful of studies that have performed genuine blind testing on wells or fields excluded from model training have found that no single method, empirical or machine-learning-based, generalizes reliably even within a single formation, a finding with direct practical implications for engineers deciding whether to adopt a published method for their own field rather than validating it against their own data first. The growing reliance on inherently opaque deep-learning architectures, set against reservoir engineering's traditionally high bar for physically justified, explainable decision-making, further suggests that interpretability will remain a live methodological concern rather than a solved problem, even as model accuracy continues to improve.

Taken together, these findings suggest that AI/ML methods in reservoir engineering have reached a stage of documented, case-by-case effectiveness without yet having reached the methodological standardization - in benchmark datasets, evaluation protocols and interpretability practice - that would allow the field's practitioners to adopt these methods with the same routine confidence long afforded to conventional physics-based reservoir engineering tools.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable - this is a review article; no new data were generated. All datasets discussed are cited to their original sources.

Conflicts of interest

The author declares no conflicts of interest.

Funding

This review received no external funding.

Authors' contributions

Not applicable.

Acknowledgements

Not applicable.

References

1. Zhao Y, Luo R, Li L, et al. A review on optimization algorithms and surrogate models for reservoir automatic history matching. *Geoenergy Sci Eng* 2024;233:212554.
2. Zhou W, Liu C, Liu Y, et al. Machine Learning in Reservoir Engineering: A Review. *Processes* 2024;12(6):1219.
3. Jing T, Lin K, Zhao W, et al. Applications and Challenges of Machine Learning in Enhanced Oil Recovery: A Review. *Energy Fuels* 2026;40(11):5395.
4. Alagoz E, Ozkan Y, Tang H, et al. AI Applications in Drilling Engineering. In: Temizel C, Tutun S, Canbaz CH, et al. *Artificial Intelligence in the Energy Industry: Theory, Case Studies and Applications*. Boca Raton: CRC Press 2026:160-234.
5. Alagoz E, Tang H, Ozdaban F, et al. AI Applications in Well Completion, Production and Stimulation. In: Temizel C, Tutun S, Canbaz CH, et al. *Artificial Intelligence in the Energy Industry: Theory, Case Studies and Applications*. Boca Raton: CRC Press 2026:291-354.
6. Alagoz E, Ozdaban F, Gozel S, et al. AI Applications in Facilities, Pipelines and Metering. *Artificial Intelligence in the Energy Industry: Theory, Case Studies and Applications*. Boca Raton: CRC Press 2026:416-477.
7. Temizel C, Tutun S, Canbaz CH, et al. AI Applications in Renewable Resources. In: Temizel C, Tutun S, Canbaz CH, et al. *Artificial Intelligence in the Energy Industry: Theory, Case Studies and Applications*. Boca Raton: CRC Press 2026:624-690.
8. Hassija V, Chamola V, Mahapatra A, et al. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cogn Comput* 2024;16(1):45-74.
9. Breiman L. Random forests. *Mach Learn* 2001;45(1):5-32.
10. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Paper presented at: 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2016.
11. Vapnik VN. *The Nature of Statistical Learning Theory*. New York: Springer 1995.
12. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735-1780.
13. Gu S, Wang J, Xue L, et al. Deep-learning-based production decline curve analysis in the gas reservoir through sequence learning models. *Comput Model Eng Sci* 2022;131(3):1579-1599.

14. Raissi M, Perdikaris P, Karniadakis GE. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 2019;378:686-707.
15. Han J-X, Xue L, Wei Y-S, et al. Physics-informed neural network-based petroleum reservoir simulation with sparse data using domain decomposition. *Pet Sci* 2023;20(6):3450-3460.
16. Narayan S, Konka S, Chandra A, et al. Accuracy assessment of various supervised machine learning algorithms in litho-facies classification from seismic data in the Penobscot field, Scotian Basin. *Front Earth Sci* 2023;11:1150954.
17. Nguyen AH, Nguyen K, Mai N. Classifying reservoir facies using attention-based residual neural networks. *PeerJ Comput Sci* 2025.
18. Xie Y, Zhu C, Zhou W, et al. Evaluation of machine learning methods for formation lithology identification: a comparison of tuning processes and model performances. *J Pet Sci Eng* 2018;160:182-193.
19. Li Z, Kang Y, Feng D, et al. Semi-supervised learning for lithology identification using Laplacian support vector machine. *J Pet Sci Eng* 2020;195:107510.
20. Antariksa G, Muammar R, Lee J. Performance evaluation of machine learning-based classification with rock-physics analysis of geological lithofacies in Tarakan Basin, Indonesia. *J Pet Sci Eng* 2022;208:109250.
21. Yang X, Dindoruk B, Lu L. A comparative analysis of bubble point pressure prediction using advanced machine learning algorithms and classical correlations. *J Pet Sci Eng* 2020;185:106598.
22. Anemangely M, Ramezanzadeh A, Amiri H, et al. Machine learning technique for the prediction of shear wave velocity using petrophysical logs. *J Pet Sci Eng* 2019;174:306-327.
23. Khan MR, Tariq Z, Abdurraheem A. Machine Learning Derived Correlation to Determine Water Saturation in Complex Lithologies. Paper presented at: SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition 2018.
24. Alqahtani N, Alzubaidi F, Armstrong RT, et al. Machine learning for predicting properties of porous media from 2D X-ray images. *J Pet Sci Eng* 2020;184:106514.
25. Cunha A, Pochet A, Lopes H, et al. Seismic fault detection in real data using transfer learning from a convolutional neural network pre-trained with synthetic seismic data. *Comput Geosci* 2020;135:104344.
26. Shahkarami A, Mohaghegh S. Applications of smart proxies for subsurface modeling. *Pet Explor Dev* 2020;47(2):400-412.
27. Tang M, Liu Y, Durlafsky LJ. A deep-learning-based surrogate model for data assimilation in dynamic subsurface flow problems. *J Comput Phys* 2020;413:109456.
28. Tang M, Liu Y, Durlafsky LJ. History Matching Complex 3D Systems Using Deep-Learning-Based Surrogate Flow Modeling and CNN-PCA Geological Parameterization. Paper presented at: SPE Reservoir Simulation Conference 2021.
29. Alakeely A, Horne RN. Simulating the Behavior of Reservoirs with Convolutional and Recurrent Neural Networks. *SPE Res Eval Eng* 2020;23:992-1005.
30. Cavalcante CCB, Maschio C, Schiozer D, et al. A stochastic learning-from-data approach to the history-matching problem. *Eng Appl Artif Intell* 2020;94:103767.
31. Chai Z, Nwachukwu A, Zagayevskiy Y, et al. An integrated closed-loop solution to assisted history matching and field optimization with machine learning techniques. *J Pet Sci Eng* 2021;198:108204.
32. Kim J, Yang H, Choe J. Robust optimization of the locations and types of multiple wells using CNN based proxy models. *J Pet Sci Eng* 2020;193:107424.
33. Wang Y, Alpak F, Gao G, et al. An Efficient Bi-Objective Optimization Workflow Using the Distributed Quasi-Newton Method and Its Application to Well-Location Optimization. *SPE J* 2022;27:364-380.
34. Li D, Liu X, Zha W, et al. Automatic well test interpretation based on convolutional neural network for a radial composite reservoir. *Pet Explor Dev* 2020;47(3):623-631.
35. Arps JJ. Analysis of decline curves. *Trans AIME* 1945;160(01):228-247.
36. Duong AN. Rate-decline analysis for fracture-dominated shale reservoirs. *SPE Res Eval Eng* 2011;14(03):377-387.
37. Clark AJ, Lake LW, Patzek TW. Production Forecasting with Logistic Growth Models. Paper presented at: SPE Annual Technical Conference and Exhibition 2011.
38. Al-Fattah SM, Startzman RA. Predicting Natural Gas Production Using Artificial Neural Network. Paper presented at: SPE Hydrocarbon Economics and Evaluation Symposium 2001.
39. Sagheer A, Kotb M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* 2019;323:203-213.
40. Ghahfarokhi PK, Carr T, Bhattacharya S, et al. A Fiber-Optic Assisted Multilayer Perceptron Reservoir Production Modeling: A Machine Learning Approach in Prediction of Gas Production from the Marcellus Shale. Paper presented at: SPE/AAPG/SEG Unconventional Resources Technology Conference 2018.
41. Sun J, Ma X, Kazi M. Comparison of Decline Curve Analysis (DCA) With Recursive Neural Networks (RNN) for Production Forecast of Multiple Wells. Paper presented at: SPE Western Regional Meeting 2018.
42. Panja P, Velasco R, Pathak M, et al. Application of artificial intelligence to forecast hydrocarbon production from shales. *Petroleum* 2018;4(1):75-89.
43. Ning Y, Kazemi H, Tahmasebi P. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM and Prophet. *Comput Geosci* 2022;164:105126.
44. Tadjer A, Hong A, Bratvold RB. Machine learning based decline curve analysis for short-term oil production forecast. *Energy Explor Exploit* 2021;39(5):1747-1769.
45. Laalam A, Khalifa H, Ouadi H, et al. Evaluation of Empirical Correlations and Time Series Models for the Prediction and Forecast of Unconventional Wells Production in Wolfcamp a Formation. Paper presented at: Unconventional Resources Technology Conference 2024.
46. Laalam A, Tomomewo OS, Khalifa H, et al. Comparative Analysis Between Empirical Correlations and Time Series Models for the Prediction and Forecasting of Unconventional Bakken Wells Production. Paper presented at: Asia Pacific Unconventional Resources Symposium 2023.
47. Matoorian R, Malaieri M, Shor R, et al. Introducing a New Hybrid Data-Physics Architecture for Production Forecasting in Unconventional Wells. Paper presented at: SPE Canadian Energy Technology Conference and Exhibition 2024.
48. Chen G, Tian H, Xiao T, et al. Time series forecasting of oil production in Enhanced Oil Recovery system based on a novel CNN-GRU neural network. *Geoenergy Sci Eng* 2024;233:212528.
49. Du J, Zheng J, Liang Y, et al. A deep learning-based approach for predicting oil production: A case study in the United States. *Energy* 2024;288:129688.
50. Su S, Zhang N, Wang P, et al. Investigation and Optimization of EOR Screening by Implementing Machine Learning Algorithms. *Appl Sci* 2023;13(22):12267.
51. Sun Q, Ertekin T. Screening and optimization of polymer flooding projects using artificial-neural-network (ANN) based proxies. *J Pet Sci Eng* 2020;185:106617.
52. Zhong Z, Sun AY, Wang Y, et al. Predicting field production rates for waterflooding using a machine learning-based proxy model. *J Pet Sci Eng* 2020;194:107574.

53. Negash BM, Yaw AD. Artificial neural network-based production forecasting for a hydrocarbon reservoir under water injection. *Pet Explor Dev* 2020;47(2):383-392.
54. Karacan CÖ. A fuzzy logic approach for estimating recovery factors of miscible CO₂-EOR projects in the United States. *J Pet Sci Eng* 2020;184:106533.
55. You J, Ampomah W, Tu J, et al. Optimization of Water-Alternating-CO₂ Injection Field Operations Using a Machine-Learning-Assisted Workflow. *SPE Res Eval Eng* 2022;25:214-231.
56. Pirizadeh M, Alemohammad N, Manthouri M, et al. A new machine learning ensemble model for class imbalance problem of screening enhanced oil recovery methods. *J Pet Sci Eng* 2021;198:108214.
57. Moeini R, Babaei M. Hybrid SVM-CIPSO methods for optimal operation of reservoir considering unknown future condition. *Appl Soft Comput* 2020;95:106572.
58. Alagoz E, Cheng RT, Al Krmagi M, et al. AI Applications in the Fourth Industry Revolution. In: Temizel C, Tutun S, Canbaz CH, et al. *Artificial Intelligence in the Energy Industry: Theory, Case Studies and Applications*. Boca Raton: CRC Press 2026:558-623.