







RESEARCH PAPER

Prediction by machine learning in nanoparticles-based enhanced oil recovery

Pavan Patel ^{1,*,\ddagger}, Saroj R. Yadav ^{1,\ddagger}, Mohamed F. El-Amin ^{2,\ddagger} and Mustafa Yıldız ^{3,\ddagger}

¹Department of Mathematics, Sardar Vallabhbhai National Institute of Technology, Surat, Gujarat, India, ²Energy Research Lab., College of Engineering, Effat University, Jeddah, Saudi Arabia,

³Department of Mathematics, Bartın University, Bartın 74100, Türkiye

*Corresponding Author

\ddaggerpavanpatel704@gmail.com (Pavan Patel); sry@amhd.svnit.ac.in (Saroj R. Yadav); momousa@effatuniversity.edu.sa (Mohamed F. El-Amin); myildiz@bartin.edu.tr (Mustafa Yıldız)

Abstract

Nanotechnology is on the brink of transforming numerous industrial sectors, and the petroleum industry stands as a front-runner in embracing these revolutionary advancements. In recent years, a growing interest has occurred in leveraging nanotechnology within the petroleum industry. Extensive research studies on nano-enhanced oil recovery (nano-EOR) have consistently delivered promising outcomes, underscoring its potential to elevate oil production substantially. However, a notable challenge persists within this domain due to the limited data availability concerning nanoparticle transport in porous media. This paper uses machine learning techniques to predict nanoparticle transport in porous media. This study uses the finite difference method to generate simulated datasets from a modified linear adsorption model. These simulated datasets are used to train machine learning models for prediction by considering artificial neural network (ANNs), decision tree (DT), and random forest (RF). We achieve mean squared values for ANN as 0.0478 (training), 0.0496 (testing), 0.0509 (validation), and R-squared values as 0.9798 (training), 0.9780 (testing), 0.9773 (validation), and for DT and RF mean squared values are 0.014683, 0.009807, and R squared values are 0.928775, 0.952425.

Keywords: Enhanced oil recovery; nanoparticles; machine learning; fluid flow

AMS 2020 Classification: 35Q35; 68T07

1 Introduction

Enhanced oil recovery (EOR) is a technique employed to augment the production of hydrocarbons from a well. The selection of an appropriate EOR method is contingent upon the data amassed

during the reservoir assessment phase. Nanoparticles (NPs) are utilized in EOR because a significant portion, approximately 60% to 70% of the hydrocarbons in most oil reservoirs remain unrecovered through primary and secondary recovery methods. Nanometer-sized particles can migrate, for instance, through the process of diffusion, into the confined pore throats within reservoir rock formations. Within these narrow spaces, these particles can engage in various physical and chemical interactions with the components of the reservoir fluid-rock system. These interactions play a pivotal role in altering the multiphase flow parameters within the reservoir. Ultimately, this phenomenon contributes significantly to enhancing oil recovery from the reservoirs, leading to more efficient and effective extraction processes [1, 2]. Silica-based NPs stand out as very promising materials in a variety of applications, owing to the ease of manufacture and the diversity in designing their surface features [3–5].

This paper employs numerical modeling, specifically based on filtration theory, to generate datasets essential for the application of machine learning techniques aimed at predicting nanoparticle transport within porous media. The motivation behind this approach is the scarcity of publicly available datasets concerning nanoparticle transport in porous media, largely due to the confidential nature of most datasets from petroleum engineering companies. To address this data gap, our study generates simulated datasets using a mathematical continuum model. This synthetic dataset is subsequently utilized as input for training machine learning models, with a specific focus on predicting both nanoparticle concentration and pore volume. The machine learning techniques utilized for these predictions encompass Artificial neural network (ANN), Decision tree (DT), and Random forest (RF) algorithms.

The paper is organized as follows: **Section 2** presents a comprehensive overview of EOR using nanoparticles. **Section 3** outlines our research methodology, while **Section 4** introduces the mathematical model employed in this study. **Section 5** introduces Machine learning modeling. Moving on to **Section 6**, we showcase the results of our research and engage in discussions. Finally, **Section 7** summarizes our findings and presents our conclusion and future research scope.

2 Background

Enhanced oil recovery with nanoparticles

Traditional oil extraction techniques, including primary and secondary methods, typically recover around one-third of the initial oil reserves present in the reservoir. Global reserves are estimated to be 1.5 trillion barrels. Using a value of one-third of 1.5 trillion barrels, the residual oil after conventional recovery is predicted to be around 1.0 trillion barrels. Recognizing the significant untapped potential in remaining oil reserves, various EOR techniques, such as thermal recovery, chemical flooding, or gas flooding, have been developed to target these vast and underutilized resources [6, 7]. Finding a cost-effective, environmentally friendly, and effective way to extract the residual oil after primary and secondary recovery is still difficult. In addition, present tertiary procedures are reliant on crude oil prices. In order to develop sustainable, economical, cost-effective, efficient, and ecologically friendly solutions, further research is required. Additionally, oil corporations are constantly under pressure and searching for innovative methods to recover trapped oil, which has accounted for a significant amount of the overall cost of oil recovery.

Due to the ongoing rise in global energy demand, technology for locating hydrocarbon sources or improving oil recovery must be developed [8]. Numerous research on the use of NPs in EOR has been undertaken. Using NPS suspensions for EOR has numerous significant advantages. These benefits include strong stability, which is mostly due to the supremacy of surface forces over gravity, allowing for constant performance. Furthermore, the properties of NPs are intrinsically controlled by their size and form, allowing for easy tailoring during the production process. The chemical

properties of NPs are strongly related to their surface coating, allowing for simple customization from hydrophilic to hydrophobic, depending on application needs. Notably, approximately 99.8% of silica NPs consist of silicon dioxide, a prevalent constituent in sandstone, rendering their use environmentally friendly. Moreover, the cost-effectiveness of NPs makes them a more economical option compared to various other chemical EOR methods.

Additionally, NPs are effective at EOR through a variety of mechanisms, including the lowering of interfacial tension, regulating mobility, and changing wettability. Laboratory tests demonstrate that by using NPs, rock wettability may be changed from strongly oil-wet to strongly water-wet, considerably increasing oil recovery [9–11].

Nanoparticles transport in porous media

Lecoanet et al. [12] conducted systematic studies on the transport of nanomaterials to assess their mobility in glass-bead columns saturated with water. These nanomaterials have sizes ranging from 1.2 nm to 303 nm. Additionally, Lecoanet et al. [13] assessed the mobility of fullerene and oxide NPs within porous media under varying flow rates. Their findings indicated that, for particles with larger diameters (>100 nm), a slower flow rate led to increased retention, characterized by delayed breakthrough and a reduced plateau value, while other flow conditions remained constant. Jeong and Kim [14] explored copper oxide NPs transport in a two-dimensional porous medium. They observed the aggregation of NPs within the pores and identified that nanoparticle deposition and aggregation were notably influenced by both flow velocity and the surfactant content present in the nanoparticle dispersion.

Wasan and Nikolov's [15] research delved into the study of how nanofluids, when combined with surfactants, behaved when applied to a solid surface, specifically focusing on their spreading characteristics. The most prevalent non-toxic inorganic substance is silica NPs, which are also less expensive to produce than other NPs [16, 17]. Ju et al. [18] divided polysilicon nanomaterials into three categories based on their wettability behavior: NWP, HLP, and LHP. Additionally, they claimed that polysilicon NPs could alter the wettability of porous surfaces when adsorbed on them. Furthermore, silica NPs show good thermal stability when heated to 650°C, as demonstrated by infrared spectrum, x-ray diffraction, and SEM examination, making them appropriate for high-temperature reservoirs [19]. Ju and Fan [20] established a model for the transport of NPs within a two-phase flow in porous media, utilizing the framework derived from the colloid model of fine particle transport in two-phase flow within porous media [21]. El-Amin et al. [22–25] have conducted simulations and modeling focused on the transport of NPs in connection with two-phase flow within porous media.

Artificial intelligence (AI) technologies are becoming more and more widespread because of their strong generalization potential and rapid reaction [26]. Machine learning shows promising potential in augmenting and improving traditional reservoir engineering methods across a broad spectrum of reservoir engineering challenges [27]. ML models discover relationships between inputs and outputs. ML may address problems such as regression, clustering, filtering, classification, and forecasting. ANN, SVM, RF, GBR, and DT are examples of machine-learning approaches. Traditional mathematical models employed in the simulation of reservoir fields are indispensable for understanding and predicting oil recovery processes. Nevertheless, they come with inherent complexities and demand substantial computational resources, leading to extended processing times. The intricate nature of these models can pose significant challenges in terms of both their development and execution. As such, there is a growing need for innovative approaches, including machine learning and data-driven techniques, to streamline and enhance the efficiency of reservoir field simulations, potentially reducing computational burdens while maintaining

accuracy and reliability. ML petroleum researchers frequently confront the challenge of globalizing their models to yield more broad correlations outside the test data set. ML applications have constraints and problems that hamper globalization, including overfitting, coincidence, excessive training, uninterpretable findings, and bias. Additionally, these models demand extensive data, which may not always be available. All these issues are confirmed by many studies [28–31].

Subasi et al. [32] created a machine learning model for predicting reservoir permeability based on well-log data using SGB regression. Their research puts many ML techniques to the test, including RF, ANN, KNN, SVM, and SGB. They discovered that SGB outperformed other evaluated models in numerous evaluation metrics tests, including accuracy and root mean squared error. In a study conducted by J. Y. Lee et al. [33], an ANN was employed to train on data from 230 successful EOR projects. The initial objective of this endeavor was to develop a predictive model capable of determining the most suitable EOR method for a candidate reservoir. By utilizing data from previous successful projects, the ANN was able to learn and identify patterns and correlations that could inform EOR method selection. This innovative approach holds significant promise in aiding decision-makers within the oil and gas industry in optimizing EOR strategies for improved reservoir performance. Irfan and Shafie [34] solved and simulated fluid flow problems using deep learning and ANN. Alwated and El-Amin [35] used ML techniques for the prediction of NPs transport in porous media. El-Amin and Alwated [36] used ML techniques for the prediction of NPs transport with the two-phase flow in porous media.

3 Research methodology

The research involved a series of distinct steps: First, developing a mathematical model for simulating NPs transport in porous media. Second, solving the model using FDM and generating simulated datasets for ML. Third, specific ML techniques, such as ANN, DT, and RF, should be implemented to predict NPs' transport in porous media. Performance metrics like MSE and R^2 were used to evaluate model effectiveness. These actions collectively constituted the research process. The strategy used to perform this research is summarized in a flow chart in Figure 1.

4 Mathematical model

Before describing the model, some assumptions are required.

- Homogeneous porous medium.
- Incompressible fluid and medium.
- Constant rate at isothermal state.
- Dispersion occurs in a parallel direction.

Advection-dispersion equation (ADE) is widely used to describe particle flow in porous media.

$$\frac{\partial c}{\partial t} = D \frac{\partial^2 c}{\partial x^2} - v \frac{\partial c}{\partial x}, \quad (1)$$

where c represents the concentration of the particle, and D represents the dispersion coefficient. $D \frac{\partial^2 c}{\partial x^2}$ represents the dispersion term and $v \frac{\partial c}{\partial x}$ represents the advection term.

The term advection refers to particle transport driven by the bulk motion of the fluid, dependent on velocity and concentration. In the context of NPs flow, advection stands out as the primary driving force behind particle motion. Dispersion, which characterizes the propensity of particles to disperse and mix due to molecular diffusion and velocity gradients, is the second essential process to take into account. Dispersion encompasses both diffusion and mechanical dispersion.

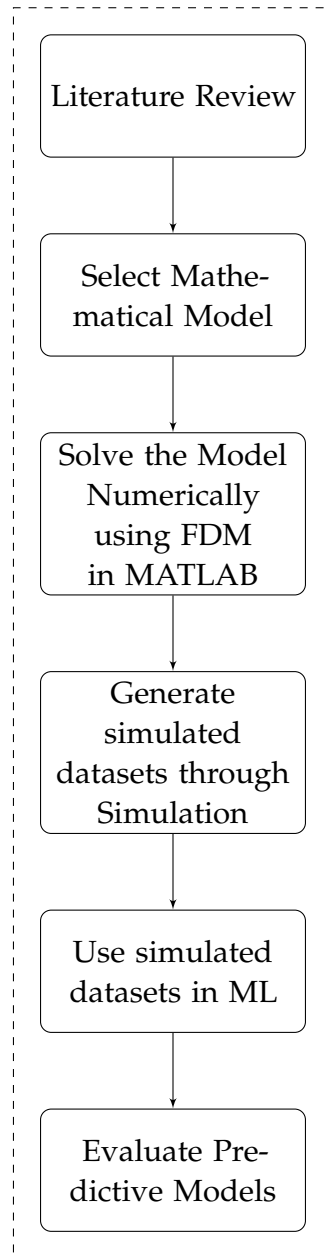


Figure 1. Flowchart of research methodology

Additionally, in certain scenarios, particle-medium surface interactions may occur, necessitating their inclusion in the equation.

Since there is always a reaction between NPs and the solid pore surface, the expression above Eq. (1) is only applicable for the description of passive tracers and fails for modeling nanoparticle flow because deposition will take place [37]. The reaction term is, therefore, included to account for the binding of particles to the grain surface. Reaction term is described as $\frac{\rho_b}{\phi} \frac{\partial s}{\partial t}$, Where ϕ is porosity, ρ_b is the bulk density, and s is the attachment concentration on solid surface.

Now, the ADE equation is

$$\frac{\partial c}{\partial t} + \frac{\rho_b}{\phi} \frac{\partial s}{\partial t} = D \frac{\partial^2 c}{\partial x^2} - v \frac{\partial c}{\partial x}. \quad (2)$$

Agista [38] proposed a modified linear adsorption model. They assume that s is linear to c means

$$s = k_c c \eta, \quad (3)$$

$$\frac{\rho_b}{\phi} \frac{\partial s}{\partial t} = \frac{\rho_b}{\phi} \frac{\partial (k_c c \eta)}{\partial t}, \quad (4)$$

where c is concentration of particle, η is the collector efficiency, and k_c is concentration distribution coefficient. By combining Eqs. (3) and (4) with Eq. (2) Hence, the final equation is written as

$$\frac{\partial c}{\partial t} + \frac{\rho_b}{\phi} \frac{\partial (k_c c \eta)}{\partial t} = D \frac{\partial^2 c}{\partial x^2} - v \frac{\partial c}{\partial x}. \quad (5)$$

The linear expression assumes that the equilibrium between the adsorbed species concentration on the solid and the dispersed concentration on the surrounding fluid is attained in a short period. This suggests that any modification in the injected concentration will have an immediate and proportional impact on the concentration of adsorbed species. As a result, adsorption is assumed to be a reversible process. By implementing a sufficient post-flush step ($c = 0$), it becomes possible to completely remove all adsorbed particles, thereby reducing the adsorbed concentration to zero ($s = 0$).

The ICs :

$$c(x, 0) = 0, s(x, 0) = 0, \quad \text{for } 0 \leq x \leq L. \quad (6)$$

The BCs are :

$$c(0, t) = \begin{cases} c_0, & \text{for } 0 \leq t \leq t_i, \\ 0, & \text{for } t > t_i, \end{cases} \quad (7)$$

$$c_x(L, t) = 0, \quad \text{for } t > 0.$$

5 Machine learning modeling

Artificial neural network

Perceptrons can only categorize instances that are linearly separable. Input instances are linearly separable if they can be divided into categories by a straight line or plane. The perceptron will then determine the answer. If the cases are non-linear, Separate learning will never be able to classify all situations accurately. Artificial neural networks have been developed to address this issue. [39] Zang presented an overview of existing work in ANN. ANN is an ML model that can recognize and classify complex patterns [40]. ANN is motivated by how human brains function and want to mimic that action. Like human brains attempt to predict patterns, ANN function as a model; they employ neurons, linked building blocks that resemble the brain's cells. These neurons get data, run it through specific rules, and produce something. Three layers make up the fundamental structure of an ANN: the input layer, hidden layers, and output layer.

Input layer nodes operate similarly to receivers in that they do not change the data they receive. Multiple nodes in the hidden layers receive the same value from them. Alternatively, they replicate the input data and deliver it to the hidden nodes for processing. Hidden layers perform considerable input value adjustments in neural networks. Either they directly access the input layer or other hidden nodes for information. Hidden layers are linked to output nodes or other hidden layers. These layers use a weighted connection approach to manage the data. Hidden layers in a neural network perform considerable input value adjustments. Either other hidden nodes or the input layer directly provides them with information. Hidden layers are linked to other hidden layers or output nodes. These layers use a system of weighted connections to manage the data. The process is as follows: Each input value is multiplied by a weight, which is a predetermined number. The weighted inputs are combined to produce a single number as the last step. By using this technique, the hidden layers can carry out complex calculations and extract important information from the incoming input. The number of hidden layers that neural networks use will depend on how challenging the problem that they are trying to solve. The output layer receives data from either the hidden or input layers. Its principal function is to return an output value that represents the expected response variable.

ANN applications may be assessed based on data analysis characteristics, including accuracy, processing speed, delay, performance, fault tolerance, volume, scalability, and convergence [41]. ANNs offer high-speed processing and parallel implementation, highlighting the need for further study in this field of research [42]. ANN is effective for resolving issues with text data, image data, and tabular data. The learning rate, which can range between 0 and 1, influences how rapidly the network learns from data. A greater learning rate means that the network learns faster, whereas a lower learning rate suggests that the network learns more slowly.

Decision tree

Murthy [43] offered an overview of work on DT and a sample of their utility for both newbies and experts in the field of machine learning. Furthermore, Quinlan [44] suggests that decision trees may be transformed into rules by defining unique rules for each path from root to leaf. Rules can be derived directly from training data using rule-based algorithms. Furnkranz [45] presented a comprehensive survey of rule-based techniques. DT is a statistical model that determines a target value based on binary rules. It comprises branches, nodes, and leaves and is a straightforward yet effective paradigm. DT is frequently used in machine learning to solve classification and regression issues [46]. They may also be seen as a series of decision-making steps, where each node stands for a particular set of rules. The evaluation moves forward by selecting one of the

Table 1. Evaluation metrics

MSE	$\frac{1}{n} \sum_{k=1}^n (\text{actual}_k - \text{predicted}_k)^2$
R ²	$1 - \frac{\text{MSE}_{\text{model}}}{\text{MSE}_{\text{base}}}$

Table 2. Simulation parameters [38, 47]

v_p (cc)	$14.8 * 10^{-6}$
ϕ (%)	47.3
d_p (nm)	100
c_i (wt%)	0.1
v (m/s)	0.003093
ρ_b (kg/m ³)	$1.28 * 10^3$
q (m ³ /s)	$8.33 * 10^{-6}$
D (m ² /s)	$2.1 * 10^{-6}$
η	0.00244
k_c	0.135

two branches based on the result of the rule, starting at the root node. The final leaf, which often represents the desired value, is reached after continuing this procedure. DT provides a simple and understandable issue-solving method in many fields.

Random forest

Ho [48] suggested the random-subspace approach, which was later formalized as the random forest by Breiman [49]. The popular supervised machine learning method RF is used for regression and classification applications. Multiple DTs are combined using ensemble learning on various dataset subsets. It improves accuracy and prevents overfitting by averaging or majority voting the predictions of individual trees. Its key strength is the algorithm's capacity to manage complicated information and make reliable predictions. The accuracy of RF increases with the number of trees by identifying more varied patterns. RF is well-known for its adaptability and consistent performance in various domains and is widely utilized in finance, healthcare, Enhanced oil recovery, and natural language processing.

Evaluation metrics

Mean squared error (MSE), and R squared (R²) are evaluation metrics that are frequently used to evaluate machine learning models. These metrics offer insights into model performance by measuring the average squared difference and standard deviation of residuals.

6 Results and discussion

Simulation parameter

Core-flooding data [47] are utilized as the input for a modified linear model in this section. Laboratory studies involved the injection of iron oxide (Experiment 92) into a saturated Boise sand pack medium and other data taken from [38]. Experiment (92) explored the sequestration of iron oxide NPs under conditions of increased flow rates and reduced particle concentrations. Simulation parameters are detailed in Table 2.

The Eq. (5), together with the initial condition (6) and boundary condition (7), is solved numerically using the FDM and implemented within the MATLAB environment. Figure 2 shows an example of running the numerical method and plotting NPs concentration versus the pore volume.

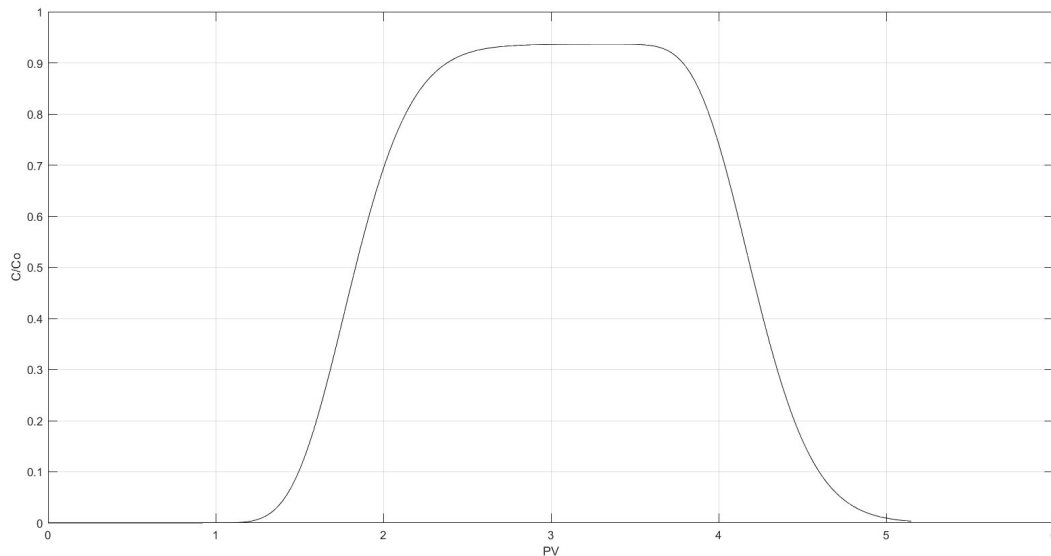


Figure 2. NPs concentration against pore volume

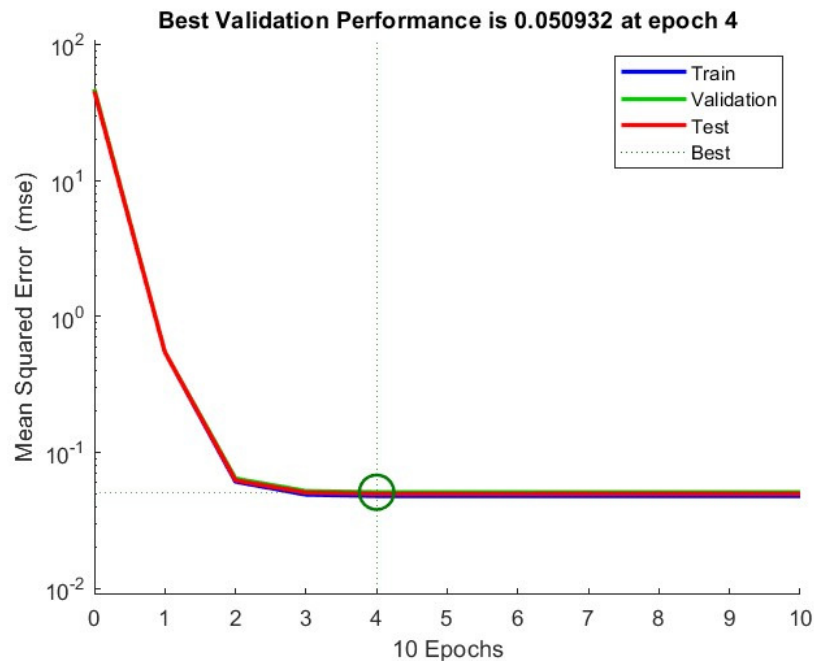
Machine learning results and discussion

The FDM approach is used to build the simulated datasets. The simulated datasets contains 10 features, including ρ_b (bulk density), D (dispersion coefficient), η (collector efficiency), k_c (distribution coefficient), v (Velocity), ϕ (porosity), t (time), x (space), c (concentration), PV (pore volume). Here c (concentration), and PV (pore volume) are selected as dependent variables for prediction, and others are selected as independent variables. **Table 3** represents The statistical information of the datasets. The dataset summary includes all instances, along with essential statistics like mean, standard deviation, minimum, quartiles, median, and maximum values for each feature. The hardware we used for training consisted of a Windows 11 Home Single Language operating system, with an AMD Ryzen 5 3500U CPU featuring 4 cores and a clock frequency of 2.10 GHz, along with 8 GB of RAM.

Simulated datasets generated from numerical simulation results are used in ML algorithms. Three different ML algorithms were used for the design prediction models, including ANN, DT, and RF. The ANN model is mainly useful in cases where complex nonlinearities are common, such as fluid dynamics and other fields, as it can learn complex non-linear dependencies. Similarly, Decision Trees and Random Forests are known for interpretability, and they reduce overfitting and deliver substantial advantages in real-world settings where actionable insights are paramount, such as surveillance systems or early feasibility studies. The data set is categorized into two parts in a ratio of 70:30, which means 70% for training and 30% for testing For DT and RF for ANN we split 30% data in testing and validation. In ANN, 10 hidden layers are employed, and the network is trained using the Levenberg-Marquardt optimization method due to fast convergence with mean square error (MSE). Tanh(x) function is considered an activation function; we trained the neural network in MATLAB (neural network toolbox). **Figure 3** represents ANN model training performance. ANN achieved the best value for the R squared method for training, testing, validation, and overall values, which are 0.9798, 0.9780, 0.9773, and 0.9792, and MSE values for training, testing, validation 0.0478, 0.0496, and 0.0509. **Figure 4** represents the Values of regression for the ANN model. **Figure 5** represents the error histogram of the ANN model. In **Figure 6**, Mu represents the gradients and step size, with a Mu parameter set at $1e^{-05}$, and the gradient value is recorded as 0.00056401. The graph illustrates a trend where both the gradient and Mu values decrease as the

Table 3. The statistical information of the dataset

	ρ_b	D	η	k_c	v	ϕ	t	x	c	PV
count	6076.0	6076.0	6076.0	6076.0	6076.0	6076.0	6076.0	6076.0	6076.0	6076.0
mean	1280.0	0.000002	0.00244	0.135	0.003093	0.473000	249.50	0.150000	0.448286	2.575000
std	0.0	0.000000	0.000000	0.000	0.000000	5.55×10^{-17}	144.08	0.086624	0.455514	1.487044
min	1280.0	0.000002	0.00244	0.135	0.003093	0.473000	0.000	0.000000	0.000000	0.000000
25%	1280.0	0.000002	0.00244	0.135	0.003093	0.473000	124.75	0.075000	0.000010	1.287500
50%	1280.0	0.000002	0.00244	0.135	0.003093	0.473000	249.50	0.150000	0.228976	2.575000
75%	1280.0	0.000002	0.00244	0.135	0.003093	0.473000	374.25	0.225000	0.995902	3.862500
max	1280.0	0.000002	0.00244	0.135	0.003093	0.473000	499.00	0.300000	1.000000	5.150000

**Figure 3.** ANN model training performance

time period progresses.

For DT and RF The dataset is split into training and testing using the train test split function from the Scikit-learn module. Four variables are produced by it: x test, y test, x train, and y train. the model is trained with the x train and y train, while the model is assessed using the x test. We used the default hyperparameters provided by the sci-kit-learn library in Python for RF (100 estimators, max features set to 1.0, and max depth set to None) and for DT (max features and max depth are none). Contrasting the anticipated value with the actual value can reveal the error and the precision of the model. DT achieved R squared, MSE, RMSE, and MAE values of 0.928775, 0.014683, 0.086180, 0.027885, and RF achieved R squared, MSE, RMSE, and MAE values of 0.952425, 0.009807, 0.070340, 0.025088 respectively. **Figure 7** represents the scatter plot of DT, and **Figure 8** represents the scatter plot of RF. In **Figure 4**, **Figure 7** and **Figure 8**, the dotted diagonal line represents the line of perfect prediction. For values ranging approximately between 1 and 5, the predictions are closely aligned with the values, as indicated by points clustering around the diagonal line. This suggests that the model is quite accurate for higher values for predictions. On the other hand, there is a noticeable clustering of points below the diagonal line for true values between 0 and around 1. The reason is that the cluster is 0 to 1 May, which is slightly outside the diagonal line because the actual values are minimal, so in this range, the performance is slightly

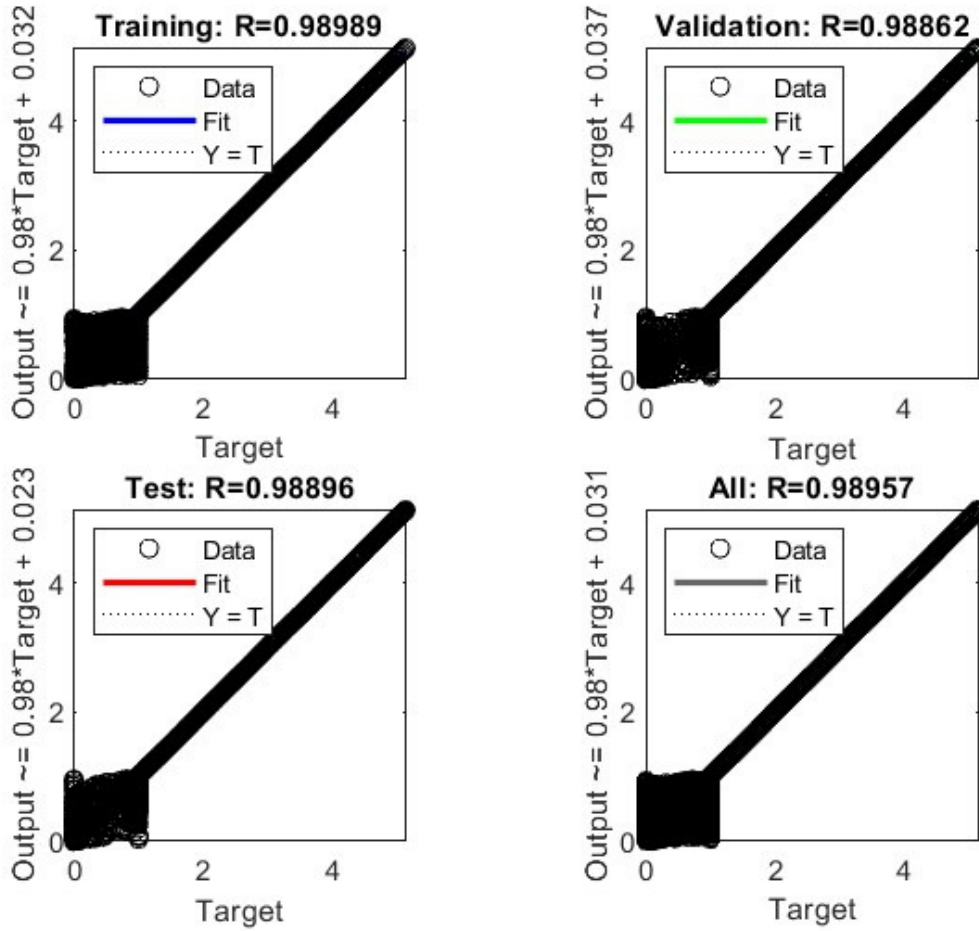


Figure 4. Values of regression for the ANN model

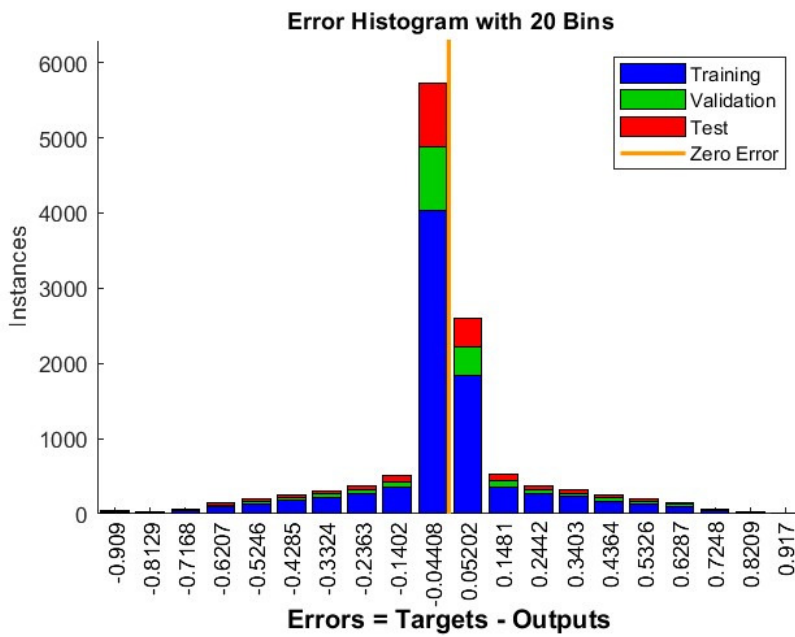


Figure 5. Error histogram

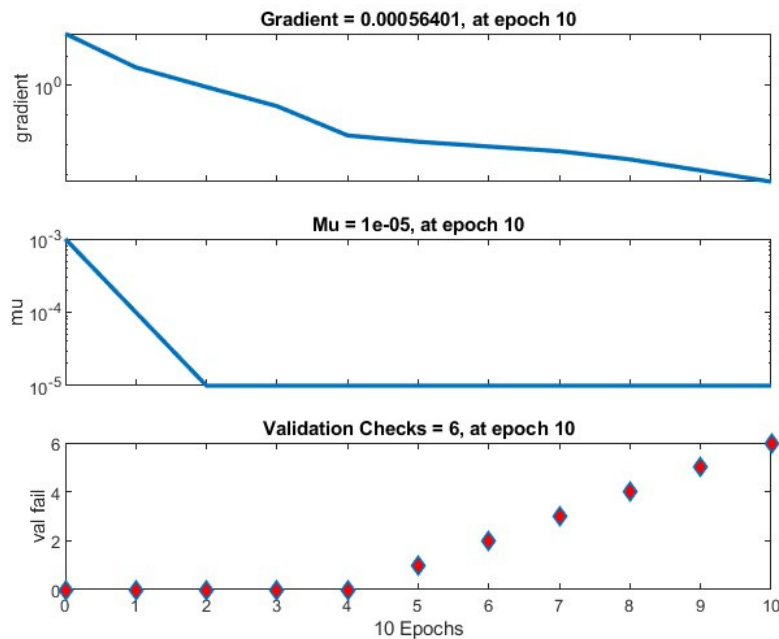


Figure 6. Transition statistics

Table 4. Evaluation metrics results for ANN, DT, and RF models

Model	R^2	MSE
ANN (Training)	0.9798	0.0478
ANN (Testing)	0.9780	0.0496
ANN (Validation)	0.9773	0.0509
DT	0.9287	0.0146
RF	0.9524	0.0098

outside the diagonal line. This indicates that the model tends to under-predict these values; the predicted values are lower than the true values. Table 4 represents the overall performance of machine learning models.

7 Conclusion and future research scope

Conclusion

Many studies on the application of nanotechnology in the petroleum sector, particularly in EOR, have yielded encouraging findings. With various advantages, nano-EOR is recommended to replace conventional chemical EOR for enhancing oil recovery efficiency: (1) NPs exhibit the remarkable ability to enhance fluid performance while requiring only minimal material usage. This efficient utilization of resources underscores the economic and environmental advantages of nanotechnology in the petroleum industry. (2) Moreover, the improvements in heat and mass transfer facilitated by nanoparticles make them a viable option for application in high-temperature conditions. This characteristic opens the door to a range of possibilities for leveraging nanotechnology in the petroleum sector, particularly in scenarios where elevated temperatures are a significant factor.

The modified Linear (ML) model is employed in our study, and its numerical solution is obtained through the Finite Difference Method. To ensure the stability and reliability of the simulation, Neumann stability analysis is also applied. This analysis is crucial in achieving stable and accurate results in our modeling and simulation process. The key reason for considering this problem as

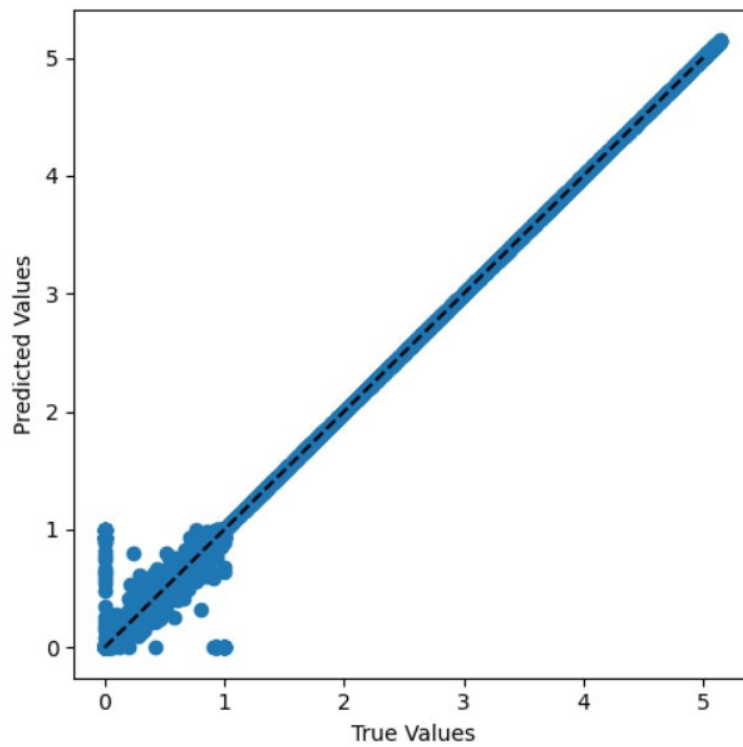


Figure 7. Scatter plot of Decision tree

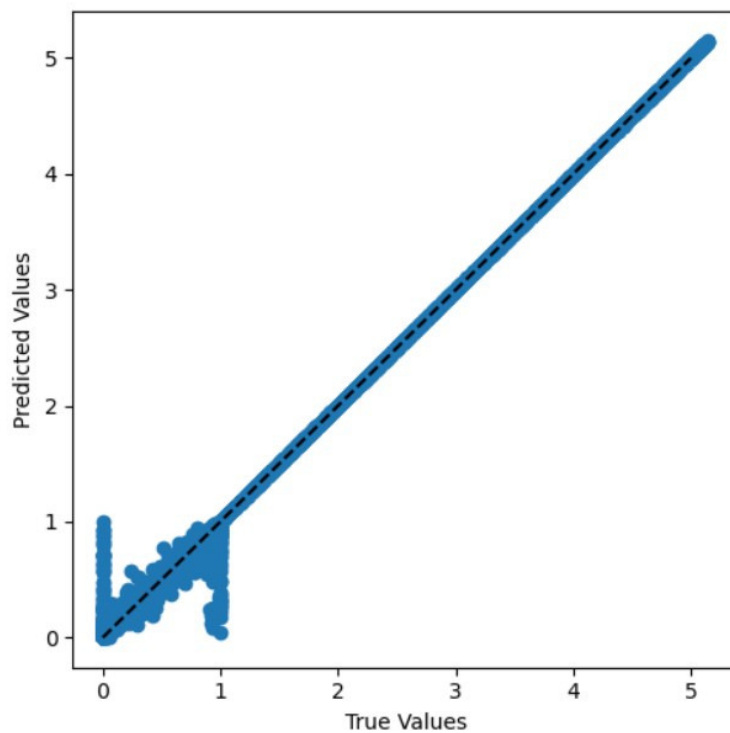


Figure 8. Scatter plot of Random forest

a machine learning (ML) task is that ML can be efficient and flexible in predicting solutions for different parameter settings. While numerical methods like FDM are known for their accuracy and robustness in solving ADE, they often require considerable computational resources, mainly when multiple solutions are ample under varying conditions. Although our FDM computational costs are relatively reasonable, in our case, FDM simulation time is approximately 20 seconds. We

explored ML as a more advanced solution that allows for quick prediction generation, which is advantageous in situations where numerous evaluations are needed. Once trained, an ML model can efficiently approximate the relationship between inputs and outputs without the necessity of redefining the system of equations.

By injecting NPs into the reservoir, reservoir, and formation parameters can be altered. The behavior of NPs in porous media can be predicted using ML. In the current study, Data is generated from mathematical models for machine learning models like ANN, DT, and RF. R-squared correlation and mean-squared error are used to evaluate the effectiveness of the predictive models. ANN model was found to have great performance.

Future research scope

Nanotechnology has increasingly gained prominence in the field of oil recovery, with its continued relevance expected in the future. Simultaneously, the era of Artificial Intelligence (AI) has ushered in new opportunities for synergy between these two fields. Due to data privacy concerns, numerous petroleum companies withhold valuable data. As a result, many models for Nano EOR have been developed empirically and theoretically. These models facilitate the generation of simulated datasets, which, in turn, serve as inputs for ML models. This predictive approach yields accurate results, thereby enabling more informed decisions that can significantly benefit the petroleum industry.

Declarations

Use of AI tools

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

List of abbreviations

ANN: Artificial Neural Network
EOR: Enhanced Oil Recovery
FDM: Finite Difference Method
HLP: Hydrophobic and Lipophilic Polysilicon
KNN: K-Nearest Neighbors
LHP: Lipophobic and Hydrophilic Polysilicon
ML: Machine Learning
NPs: Nanoparticles
NWP: Neutral-Wet Polysilicon
SGB: Stochastic Gradient Boosting
SVM: Support Vector Machine
MAE: Mean absolute error
RMSE: Root Mean Squared Error

Data availability statement

No Data associated with the manuscript.

Consent for publication

Not applicable

Conflicts of 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.

Funding

No funding was received for the paper.

Author's contributions

P.P.: Conceptualization, Methodology, Software, Writing-Original draft preparation. S.R.Y.: Supervision, Conceptualization, Methodology, Software, Writing-Original draft preparation, Writing - Review & Editing. M.F.E.: Supervision, Conceptualization, Methodology, Software, Writing-Original draft preparation, Writing - Review & Editing. M.Y.: Conceptualization, Methodology, Software, Review & Editing. All authors have read and approved the published version of the manuscript.

Acknowledgements

Not applicable

References

- [1] Bera, A. and Belhaj, H. Application of nanotechnology by means of nanoparticles and nanodispersions in oil recovery-A comprehensive review. *Journal of Natural Gas Science and Engineering*, 34, 1284-1309, (2016). [[CrossRef](#)]
- [2] Negin, C., Ali, S. and Xie, Q. Application of nanotechnology for enhancing oil recovery-A review. *Petroleum*, 2(4), 324-333, (2016). [[CrossRef](#)]
- [3] Metin, C.O., Baran, J.R. and Nguyen, Q.P. Adsorption of surface functionalized silica nanoparticles onto mineral surfaces and decane/water interface. *Journal of Nanoparticle Research*, 14, 1246, (2012). [[CrossRef](#)]
- [4] Miranda, C.R., De Lara, L.S. and Tonetto, B.C. Stability and mobility of functionalized silica nanoparticles for enhanced oil recovery applications. In Proceedings, *SPE International Oilfield Nanotechnology Conference and Exhibition*, pp. 157033, Noordwijk, The Netherlands, (2012, June). [[CrossRef](#)]
- [5] Nath, N., Chakraborty, S., Panda, P. and Pal, K. High yield silica-based emerging nanoparticles activities for hybrid catalyst applications. *Topics in Catalysis*, 65(19), 1706-1718, (2022). [[CrossRef](#)]
- [6] Bentley, R.W., Mannan, S.A. and Wheeler, S.J. Assessing the date of the global oil peak: the need to use 2P reserves. *Energy policy*, 35(12), 6364-6382, (2007). [[CrossRef](#)]
- [7] Katende, A. and Sagala, F. A critical review of low salinity water flooding: Mechanism, laboratory and field application. *Journal of Molecular Liquids*, 278, 627-649, (2019). [[CrossRef](#)]
- [8] Kong, X. and Ohadi, M.M. Applications of micro and nano technologies in the oil and gas industry-an overview of the recent progress. In Proceedings, *Abu Dhabi International Petroleum Exhibition and Conference*, pp. 138241, Abu Dhabi, UAE, (2010, November). [[CrossRef](#)]
- [9] Karimi, A., Fakhroueian, Z., Bahramian, A., Pour Khiabani, N., Darabad, J.B., Azin, R. and Arya, S. Wettability alteration in carbonates using zirconium oxide nanofluids: EOR implications. *Energy and Fuels*, 26(2), 1028-1036, (2012). [[CrossRef](#)]

-
- [10] Ehtesabi, H., Ahadian, M.M., Taghikhani, V. and Ghazanfari, M.H. Enhanced heavy oil recovery in sandstone cores using TiO_2 nanofluids. *Energy and Fuels*, 28(1), 423-430, (2014). [[CrossRef](#)]
- [11] Safari, M. Variations in wettability caused by nanoparticles. *Petroleum Science and Technology*, 32(12), 1505-1511, (2014). [[CrossRef](#)]
- [12] Lecoanet, H.F., Bottero, J.Y. and Wiesner, M.R. Laboratory assessment of the mobility of nanomaterials in porous media. *Environmental Science and Technology*, 38(19), 5164-5169, (2004). [[CrossRef](#)]
- [13] Lecoanet, H.F. and Wiesner, M.R. Velocity effects on fullerene and oxide nanoparticle deposition in porous media. *Environmental Science Technology*, 38(16), 4377-4382, (2004). [[CrossRef](#)]
- [14] Jeong, S.W. and Kim, S.D. Aggregation and transport of copper oxide nanoparticles in porous media. *Journal of Environmental Monitoring*, 11(9), 1595-1600, (2009). [[CrossRef](#)]
- [15] Wasan, D.T. and Nikolov, A.D. Spreading of nanofluids on solids. *Nature*, 423, 156-159, (2003). [[CrossRef](#)]
- [16] Hendraningrat, L. *Unlocking the Potential of Hydrophilic Nanoparticles as Novel Enhanced Oil Recovery Method: An Experimental Investigation*. Ph.D. Thesis, Norwegian University of Science and Technology - NTNU, (2015). [<https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2373540>]
- [17] Meena, J., Gupta, A., Ahuja, R., Singh, M., Bhaskar, S. and Panda, A.K. Inorganic nanoparticles for natural product delivery: A review. *Environmental Chemistry Letters*, 18, 2107-2118, (2020). [[CrossRef](#)]
- [18] Binshan, J., Shugao, D., Zhian, L., Tiangao, Z., Xiantao, S. and Xiaofeng, Q. A study of wettability and permeability change caused by adsorption of nanometer structured polysilicon on the surface of porous media. In Proceedings, *SPE Asia Pacific Oil and Gas Conference and Exhibition*, pp. 77938, Melbourne, Australia, (2002, October). [[CrossRef](#)]
- [19] Wang, L., Wang, Z., Yang, H. and Yang, G. The study of thermal stability of the SiO_2 powders with high specific surface area. *Materials Chemistry and Physics*, 57(3), 260-263, (1999). [[CrossRef](#)]
- [20] Ju, B. and Fan, T. Experimental study and mathematical model of nanoparticle transport in porous media. *Powder Technology*, 192(2), 195-202, (2009). [[CrossRef](#)]
- [21] Liu, X. and Civan, F. A multiphase mud fluid infiltration and filter cake formation model. In Proceedings, *International Symposium on Oilfield Chemistry*, pp. 607-621, Richardson, TX, United States, (1993).
- [22] El-Amin, M.F., Salama, A. and Sun, S. Modeling and simulation of nanoparticles transport in a two-phase flow in porous media. In Proceedings, *SPE International Oilfield Nanotechnology Conference and Exhibition*, pp. 154972, Noordwijk, The Netherlands, (2012, June). [[CrossRef](#)]
- [23] El-Amin, M.F., Sun, S. and Salama, A. Modeling and simulation of nanoparticle transport in multiphase flows in porous media: CO_2 sequestration. In Proceedings, *SPE Mathematical Methods in Fluid Dynamics and Simulation of Giant Oil and Gas Reservoirs*, pp. 163089, Istanbul, Turkey, (2012, September). [[CrossRef](#)]
- [24] El-Amin, M.F., Salama, A. and Sun, S. Numerical and dimensional analysis of nanoparticles transport with two-phase flow in porous media. *Journal of Petroleum Science and Engineering*, 128, 53-64, (2015). [[CrossRef](#)]
- [25] Salama, A., Negara, A., El Amin, M. and Sun, S. Numerical investigation of nanoparticles

- transport in anisotropic porous media. *Journal of Contaminant Hydrology*, 181, 114-130, (2015). [[CrossRef](#)]
- [26] Evans, S.J. How digital engineering and cross-industry knowledge transfer is reducing project execution risks in oil and gas. In Proceedings, *Offshore Technology Conference*, p. 29458, Houston, Texas, (2019, April). [[CrossRef](#)]
- [27] Anifowose, F.A., Labadin, J. and Abdulraheem, A. Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization. *Journal of Petroleum Science and Engineering*, 151, 480-487, (2017). [[CrossRef](#)]
- [28] Andrea, T.A. and Kalayeh, H. Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. *Journal of Medicinal Chemistry*, 34(9), 2824-2836, (1991).
- [29] Livingstone, D.J., Manallack, D.T. and Tetko, I.V. Data modelling with neural networks: Advantages and limitations. *Journal of Computer-Aided Molecular Design*, 11, 135-142, (1997). [[CrossRef](#)]
- [30] Weyrauch, T. and Herstatt, C. What is frugal innovation? Three defining criteria. *Journal of Frugal Innovation*, 2, 1, (2017). [[CrossRef](#)]
- [31] Hossain, M. Frugal innovation: A review and research agenda. *Journal of Cleaner Production*, 182, 926-936, (2018). [[CrossRef](#)]
- [32] Subasi, A., El-Amin, M.F., Darwich, T. and Dossary, M. Permeability prediction of petroleum reservoirs using stochastic gradient boosting regression. *Journal of Ambient Intelligence and Humanized Computing*, 13, 3555–3564, (2022). [[CrossRef](#)]
- [33] Lee, J.Y., Shin, H.J. and Lim, J.S. Selection and evaluation of enhanced oil recovery method using artificial neural network. *Geosystem Engineering*, 14(4), 157-164, (2011). [[CrossRef](#)]
- [34] Irfan, S.A. and Shafie, A. Artificial neural network modeling of nanoparticles assisted enhanced oil recovery. In *Advanced Methods for Processing and Visualizing the Renewable Energy: A New Perspective from Signal to Image Recognition* (Vol. 320, pp. 59-75). Singapore: Springer, (2021). [[CrossRef](#)]
- [35] Alwated, B. and El-Amin, M.F. Enhanced oil recovery by nanoparticles flooding: From numerical modeling improvement to machine learning prediction. *Advances in Geo-Energy Research*, 5(3), 297-317, (2021). [[CrossRef](#)]
- [36] El-Amin, M.F., Alwated, B. and Hoteit, H.A. Machine learning prediction of nanoparticle transport with two-phase flow in porous media. *Energies*, 16(2), 678, (2023). [[CrossRef](#)]
- [37] Zhang, T., Murphy, M., Yu, H., Huh, C. and Bryant, S.L. Mechanistic model for nanoparticle retention in porous media. *Transport in Porous Media*, 115, 387-406, (2016). [[CrossRef](#)]
- [38] Agista, M.N. *A Literature Review and Transport Modelling of Nanoparticles for Enhanced Oil Recovery*. Master Thesis, Department of Petroleum Technology, University of Stavanger, (2017). [<https://uis.brage.unit.no/uis-xmlui/handle/11250/2463419>]
- [39] Zhang, G.P. Neural networks for classification: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 30(4), 451-462, (2000). [[CrossRef](#)]
- [40] Del Castillo, A.A., Santoyo, E. and García-Valladares, O. A new void fraction correlation inferred from artificial neural networks for modeling two-phase flows in geothermal wells. *Computers & Geosciences*, 41, 25-39, (2012). [[CrossRef](#)]
- [41] He, H. and Garcia, E.A. Learning from imbalanced data. *IEEE Transactions on Knowledge and*

Data Engineering, 21(9), 1263-1284, (2009). [[CrossRef](#)]

- [42] Izeboudjen, N., Larbes, C. and Farah, A. A new classification approach for neural networks hardware: from standards chips to embedded systems on chip. *Artificial Intelligence Review*, 41, 491-534, (2014). [[CrossRef](#)]
- [43] Murthy, S.K. Automatic construction of decision trees from data: A multi-disciplinary survey. *Data Mining and Knowledge Discovery*, 2, 345-389, (1998). [[CrossRef](#)]
- [44] Quinlan, J.R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers: San Mateo, California, (1993).
- [45] Fürnkranz, J. Separate-and-conquer rule learning. *Artificial Intelligence Review*, 13, 3-54, (1999). [[CrossRef](#)]
- [46] Kotsiantis, S.B. Decision trees: a recent overview. *Artificial Intelligence Review*, 39, 261-283, (2013). [[CrossRef](#)]
- [47] Murphy, M.J. *Experimental Analysis of Electrostatic and Hydrodynamic Forces Affecting Nanoparticle Retention in Porous Media*. Ph.D. Thesis, Department of Petroleum Engineering, The University of Texas, (2012). [<https://repositories.lib.utexas.edu/items/b2d2bc2d-aed4-4bbb-8278-9c2a1078ed5e>]
- [48] Ho, T.K. Random decision forests. In Proceedings, *3rd International Conference on Document Analysis and Recognition*, pp. 278-282, Montreal, QC, Canada, (1995, August). [[CrossRef](#)]
- [49] Breiman, L. Random forests. *Machine Learning*, 45, 5-32, (2001). [[CrossRef](#)]

Mathematical Modelling and Numerical Simulation with Applications (MMNSA)
(<https://dergipark.org.tr/en/pub/mmnsa>)



Copyright: © 2024 by the authors. This work is licensed under a Creative Commons Attribution 4.0 (CC BY) International License. The authors retain ownership of the copyright for their article, but they allow anyone to download, reuse, reprint, modify, distribute, and/or copy articles in MMNSA, so long as the original authors and source are credited. To see the complete license contents, please visit (<http://creativecommons.org/licenses/by/4.0/>).

How to cite this article: Patel, P., Yadav, S.R., El-Amin, M.F. and Yıldız, M. (2024). Prediction by machine learning in nanoparticles-based enhanced oil recovery. *Mathematical Modelling and Numerical Simulation with Applications*, 4(4), 544-561. <https://doi.org/10.53391/mmnsa.1498986>