ISSN 2518-170X (Online)
ISSN 2224-5278 (Print)

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN, SERIES OF GEOLOGY AND TECHNICAL SCIENCES

№3
2026

ISSN 2518-170X (Online)
ISSN 2224-5278 (Print)

CENTRAL ASIAN ACADEMIC
RESEARCH CENTER

N E W S
OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN, SERIES OF GEOLOGY AND TECHNICAL SCIENCES

3 (477)

JUNE – JULY 2026

THE JOURNAL WAS FOUNDED IN 1940

PUBLISHED 6 TIMES A YEAR

ALMATY, 2026

The scientific journal News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences has been indexed in the international abstract and citation database Scopus since 2016 and demonstrates stable bibliometric performance.

The journal is also included in the Emerging Sources Citation Index (ESCI) of the Web of Science platform (Clarivate Analytics, since 2018).

Indexing in ESCI confirms the journal's compliance with international standards of scientific peer review and editorial ethics and is considered by Clarivate Analytics as part of the evaluation process for potential inclusion in the Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI), and Arts & Humanities Citation Index (AHCI).

Indexing in Scopus and Web of Science ensures high international visibility of publications, promotes citation growth, and reflects the editorial board's commitment to publishing relevant, original, and scientifically significant research in the fields of geology and technical sciences.

«Қазақстан Республикасы Ұлттық ғылым академиясының Хабарлары. Геология және техникалық ғылымдар сериясы» ғылыми журналы 2016 жылдан бастап халықаралық реферативтік және ғылымиметриялық Scopus дерекқорында индекстеледі және тұрақты библиометриялық көрсеткіштерді көрсетіп келеді.

Сонымен қатар журнал Web of Science платформасының (Clarivate Analytics, 2018) халықаралық реферативтік және наукометриялық дерекқоры Emerging Sources Citation Index (ESCI) тізіміне енгізілген.

ESCI дерекқорында индекстелуі журналдың халықаралық ғылыми рецензиялау талаптары мен редакциялық этика стандарттарына сәйкестігін растайды, сондай-ақ Clarivate Analytics компаниясы тарапынан басылымды Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) және Arts & Humanities Citation Index (AHCI) дерекқорларына енгізу қарастырылуда.

Scopus және Web of Science дерекқорларында индекстелуі жарияланымдардың халықаралық деңгейде жоғары сұранысқа ие болуын қамтамасыз етеді, олардың дәйексөз алу көрсеткіштерінің артуына ықпал етеді және редакциялық алқаның геология мен техникалық ғылымдар саласындағы өзекті, бірегей және ғылыми тұрғыдан маңызды зерттеулерді жариялауға ұмтылысын айқындайды.

Научный журнал «News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences» с 2016 года индексируется в международной реферативной и наукометрической базе данных Scopus и демонстрирует стабильные библиометрические показатели.

Журнал также включён в международную реферативную и наукометрическую базу данных Emerging Sources Citation Index (ESCI) платформы Web of Science (Clarivate Analytics, 2018).

Индексирование в ESCI подтверждает соответствие журнала международным стандартам научного рецензирования и редакционной этики, а также рассматривается компанией Clarivate Analytics в рамках дальнейшего включения издания в Science Citation Index Expanded (SCIE), Social Sciences Citation Index (SSCI) и Arts & Humanities Citation Index (AHCI).

Индексирование в Scopus и Web of Science обеспечивает высокую международную востребованность публикаций, способствует росту цитируемости и подтверждает стремление редакционной коллегии публиковать актуальные, оригинальные и научно значимые исследования в области геологии и технических наук.

News of the National Academy of Sciences of the Republic of Kazakhstan. Series of geology and technology sciences.
ISSN 2518-170X (Online),
ISSN 2224-5278 (Print)
Owner: «Central Asian Academic Research Center» LLP (Almaty).
The certificate of registration of a periodical printed publication in the Committee of information of the Ministry of Information and Communications of the Republic of Kazakhstan № KZ50VPY00121155, issued on 05.06.2025
Thematic scope: *geology, hydrogeology, geography, mining and chemical technologies of oil, gas and metals*
Periodicity: 6 times a year.
http://www.geolog-technical.kz/index.php/en/

News of the National Academy of Sciences of the Republic of Kazakhstan. Series of geology and technology sciences.
ISSN 2518-170X (Online),
ISSN 2224-5278 (Print)

Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы қ.).

Қазақстан Республикасының Ақпарат және коммуникациялар министрлігінің Ақпарат комитетінде 05.06.2025 ж. берілген № KZ50VPY00121155 мерзімдік басылым тіркеуіне қойылу туралы куәлік.

Тақырыптық бағыты: *геология, гидрогеология, география, тау-кен ісі, мұнай, газ және металдардың химиялық технологиялары*

Мерзімділігі: жылына 6 рет.

http://www.geolog-technical.kz/index.php/en/

News of the National Academy of Sciences of the Republic of Kazakhstan. Series of geology and technology sciences.
ISSN 2518-170X (Online),
ISSN 2224-5278 (Print)
Собственник: ТОО «Центрально-Азиатский академический научный центр» (г. Алматы).
Свидетельство о постановке на учет периодического печатного издания в Комитете информации Министерства информации и коммуникаций и Республики Казахстан № KZ50VPY00121155, выданное 05.06.2025 г.
Тематическая направленность: *геология, гидрогеология, география, горное дело и химические технологии нефти, газа и металлов*
Периодичность: 6 раз в год.
http://www.geolog-technical.kz/index.php/en/

CONTENTS

NEWS OF THE NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC
OF KAZAKHSTAN, SERIES OF GEOLOGY AND TECHNICAL SCIENCES
ISSN 2224–5278
Volume 3.
Number 477 (2026), 392–409

https://doi.org/10.32014/2026.2518-170X.660

UDC: 622.276.031
IRSTI: 52.47.17

©Zaurbekov K.S., Smailov S.M.*, Zaurbekov S.A., 2026.
Institute of Metallurgy and Ore Beneficiation, Satbayev University,
Almaty, Kazakhstan.
*E-mail: smailov.serzhan@mail.ru

APPLICATION OF MACHINE LEARNING FOR PREDICTING RELATIVE PERMEABILITIES IN CORE FLOODING: GLOBAL EXPERIENCE AND NUMERICAL EXPERIMENT

Zaurbekov Kadyrzhan — PhD, Associate Professor, Institute of Metallurgy and Ore Beneficiation, Satbayev University, Almaty, Kazakhstan,
E-mail: kadmen.95@mail.ru, https://orcid.org/0000-0001-7425-7448;
Smailov Serzhan — Junior Researcher, Satbayev University, Almaty, Kazakhstan,
E-mail: smailov.serzhan@mail.ru, https://orcid.org/0009-0001-0930-6837;
Zaurbekov Seitzhan — Candidate of Technical Sciences, Professor, Institute of Metallurgy and Ore Beneficiation, Satbayev University, Almaty, Kazakhstan,
E-mail: s.zaurbek@mail.ru, https://orcid.org/0000-0001-8025-0824.

Abstract. *Relevance.* The intensification of oil production and the growing volume of laboratory core flooding data create a pressing need for more efficient methods of filtration characteristic interpretation. Traditional analytical approaches are labor-intensive and limited in their capacity to capture complex nonlinear relationships between rock and fluid parameters, making the application of machine learning (ML) highly relevant in modern petroleum engineering. *Objective.* To substantiate the effectiveness of regression and ensemble ML methods for predicting water and oil relative permeabilities in core flooding experiments based on a controlled numerical experiment using a synthetic dataset. *Methods.* A numerical experiment was conducted in Python using a synthetic dataset of 1,200 samples generated on the basis of a modified Corey model, incorporating seven input parameters: porosity, absolute permeability, water saturation, oil and water viscosity, pressure drop, and a CT-based heterogeneity index. A 3% random noise component was introduced to simulate realistic measurement conditions. Two independent Random Forest regression models were trained to predict water relative permeability (krw) and oil relative permeability (kro), with performance evaluated using R^2 and RMSE metrics and feature importance assessed via permutation analysis. *Results.* Both models demonstrated high predictive accuracy: $R^2 = 0.996$ for krw and $R^2 = 0.994$ for kro, with RMSE not exceeding 0.012. The models accurately reproduced the nonlinear shape of relative permeability curves as a function of water saturation. Feature importance analysis identified water saturation, absolute permeability, and fluid viscosity as dominant controlling factors, consistent with the physical theory of two-phase filtration. *Conclusions.* The proposed approach improves the accuracy of relative permeability estimation, reduces experimental workload, and enables the construction of digital core models. The results confirm the ability of ML models to reproduce complex filtration dependencies without explicit physical equations while maintaining interpretability, substantiating the potential of integrating machine learning into laboratory core flooding workflows for reservoir simulation and digital twin development.

Keywords: machine learning (ML), core flooding, relative permeability, random forest, regression analysis, R^2, RMSE

Financing*. This study was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan. (Grant No. BR24992868).*

For citations: Zaurbekov K.S., Smailov S.M., Zaurbekov S.A. Application of Machine Learning for Predicting Relative Permeabilities in Core Flooding: Global Experience and Numerical Experiment. *News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences. 2026. No.3. Pp. 392–409. DOI: https://doi.org/10.32014/2026.2518-170X.660*

Металлургия және кен байыту институты, Satbayev University,
Алматы, Қазақстан.
*E-mail: smailov.serzhan@mail.ru

КЕРНДІ СУ АЙДАУ КЕЗІНДЕ САЛЫСТЫРМАЛЫ ӨТКІЗГІШТІКТЕРДІ БОЛЖАУДА МАШИНАЛЫҚ ОҚЫТУ: ӘЛЕМДІК ТӘЖІРИБЕ ЖӘНЕ САНДЫҚ ЭКСПЕРИМЕНТ

Заурбеков Кадыржан — PhD, қауымдастырылған профессор, Металлургия және кен байыту институты, Satbayev University, Алматы, Қазақстан,
E-mail: kadmen.95@mail.ru, https://orcid.org/0000-0001-7425-7448;
Смайлов Сержан — кіші ғылыми қызметкер, Satbayev University, Алматы, Қазақстан,
E-mail: smailov.serzhan@mail.ru, https://orcid.org/0009-0001-0930-6837;
Заурбеков Сейтжан — техника ғылымдарының кандидаты, профессор, Металлургия және кен байыту институты, Satbayev University, Алматы, Қазақстан,
E-mail: s.zaurbek@mail.ru, https://orcid.org/0000-0001-8025-0824.

Аннотация. *Өзектілігі.* Мұнай өндіруді қарқындату және кернді су айдау бойынша зертханалық деректер көлемінің өсуі фильтрациялық сипаттамаларды интерпретациялаудың тиімді әдістеріне деген қажеттілікті күшейтеді. Дәстүрлі аналитикалық тәсілдер жоғары еңбек шығынымен және жыныс пен флюид параметрлері арасындағы сызықтық емес тәуелділіктерді ескерудегі шектеулермен сипатталады, бұл машиналық оқытуды (МО) қазіргі заманғы мұнай инженериясында қолданудың өзектілігін айқындайды. *Мақсат.* Синтетикалық деректер жиынын пайдалана отырып жүргізілген бақыланатын сандық эксперимент негізінде кернді су айдау эксперименттерінде судың және мұнайдың салыстырмалы өткізгіштіктерін болжауға арналған регрессиялық және ансамбльдік МО әдістерінің тиімділігін негіздеу. *Әдістері.* Сандық эксперимент Python ортасында модификацияланған Corey моделі негізінде жасалған 1200 бақылаудан тұратын синтетикалық деректер жиынын пайдалана отырып жүргізілді. Деректер жиыны жеті кіріс параметрін қамтыды: кеуектілік, абсолютті өткізгіштік, су қанығуы, мұнай мен судың тұтқырлығы, қысым айырымы және КТ деректері бойынша біртектілік индексі. Нақты өлшеу жағдайларын имитациялау үшін 3% кездейсоқ шу енгізілді. Судың (krw) және мұнайдың (kro) салыстырмалы өткізгіштіктерін болжауға арналған екі тәуелсіз Random Forest регрессиялық моделі құрылды; сапа R^2 және RMSE метрикалары, белгілер маңыздылығы айналмалы талдау әдісі арқылы бағаланды. *Нәтижелер.* Екі модель де жоғары дәлдікті көрсетті: krw үшін R^2 = 0,996 және kro үшін R^2 = 0,994, RMSE мәндері 0,012-ден аспады. Модельдер салыстырмалы өткізгіштік қисықтарының су қанығуына тәуелді сызықтық емес пішінін дәл жаңғыртты. Белгілер маңыздылығын талдау су қанығуы, абсолютті өткізгіштік және флюид тұтқырлығының басым факторлар екенін анықтады, бұл екі фазалы фильтрация теориясымен толық сәйкес келеді. *Қорытындылар.* Ұсынылған тәсіл салыстырмалы өткізгіштікті бағалау дәлдігін арттырады, эксперименттік жұмыс көлемін азайтады және кернның цифрлық модельдерін жасауға негіз қалайды. Нәтижелер МО модельдерінің физикалық интерпретациялауды сақтай отырып күрделі фильтрациялық тәуелділіктерді жаңғырту қабілетін растайды және қабат гидродинамикалық модельдеуі үшін машиналық оқытуды зертханалық тәжірибеге интеграциялаудың жоғары әлеуетін негіздейді.

Түйін сөздер: машиналық оқыту (ML), кернді су айдау, салыстырмалы өткізгіштік, fandom forest, регрессиялық талдау, R^2, RMSE

©Заурбеков К.С., Смайлов С.М.*, Заурбеков С.А., 2026.
Институт металлургии и обогащения, Satbayev University,
Алматы, Казахстан.
*E-mail: smailov.serzhan@mail.ru

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ОТНОСИТЕЛЬНЫХ ПРОНИЦАЕМОСТЕЙ ПРИ ЗАВОДНЕНИИ КЕРНА: МИРОВОЙ ОПЫТ И ЧИСЛЕННЫЙ ЭКСПЕРИМЕНТ

Заурбеков Кадыржан — PhD, ассоциированный профессор, Институт металлургии и обогащения, Satbayev University, Алматы, Казахстан,
E-mail: kadmen.95@mail.ru, https://orcid.org/0000-0001-7425-7448;
Смайлов Сержан — младший научный сотрудник, Satbayev University, Алматы, Казахстан,
E-mail: smailov.serzhan@mail.ru, https://orcid.org/0009-0001-0930-6837;
Заурбеков Сейтжан — кандидат технических наук, профессор, Институт металлургии и обогащения, Satbayev University, Алматы, Казахстан,
E-mail: s.zaurbek@mail.ru, https://orcid.org/0000-0001-8025-0824.

Аннотация. *Актуальность.* Интенсификация добычи нефти и рост объемов лабораторных данных по заводнению керна обусловливают необходимость применения более эффективных методов интерпретации фильтрационных характеристик. Традиционные аналитические подходы отличаются высокой трудоемкостью и ограниченной способностью учитывать нелинейные зависимости между параметрами породы и флюидов, что определяет актуальность использования машинного обучения (МО) в современной нефтяной инженерии. *Цель*. Обосновать эффективность регрессионных и ансамблевых методов машинного обучения для прогнозирования относительных проницаемостей воды и нефти в экспериментах по заводнению керна на основе контролируемого численного эксперимента с использованием синтетического набора данных. *Методы*. Численный эксперимент проводился в среде Python с использованием синтетического набора данных из 1200 наблюдений, сгенерированного на основе модифицированной модели Кори. Набор данных включал семь входных параметров: пористость, абсолютную проницаемость, водонасыщенность, вязкость нефти и воды, перепад давления и индекс неоднородности по данным КТ. Для имитации реальных условий измерений был введен случайный шум на уровне 3%. Построены две независимые регрессионные модели Random Forest для прогнозирования относительной проницаемости воды (krw) и нефти (kro). Качество моделей оценивалось по метрикам R^2 и RMSE, а важность признаков - методом перестановочного анализа. *Результаты и выводы*. Обе модели продемонстрировали высокую точность: $R^2 = 0{,}996$ для krw и $R^2 = 0{,}994$ для kro при RMSE не более 0,012. Модели точно воспроизвели нелинейную форму кривых относительных проницаемостей как функции водонасыщенности.

Анализ важности признаков показал, что водонасыщенность, абсолютная проницаемость и вязкость флюидов являются доминирующими факторами, что соответствует физической теории двухфазной фильтрации. Предложенный подход повышает точность оценки относительных проницаемостей, сокращает объем экспериментальных работ и создает основу для построения цифровых моделей керна. Результаты подтверждают способность моделей машинного обучения воспроизводить сложные фильтрационные зависимости без явного использования физических уравнений при сохранении интерпретируемости, что обосновывает потенциал интеграции машинного обучения в лабораторную практику заводнения керна, гидродинамическое моделирование и разработку цифровых двойников.

Ключевые слова: машинное обучение (ML), заводнение керна, относительная проницаемость, random forest, регрессионный анализ, R^2, RMSE

Introduction. The development of core flooding technologies is accompanied by the active implementation of machine learning (ML) methods, which enable the analysis of complex nonlinear relationships and the prediction of filtration process outcomes. Traditional laboratory core flooding experiments, while being a reliable tool for studying the physicochemical properties of reservoirs, are labor-intensive, costly, and have limited scalability. In this context, machine learning becomes an effective tool for accelerating data analysis, improving the accuracy of experiment interpretation, and creating digital models of the reservoir.

Globally, machine learning is applied in various directions of laboratory research, ranging from the prediction of relative permeabilities and oil recovery factors (Agzamov et al., 2022; Alali et al., 2025; Steineder et al., 2020; Ginsburg, 2017) to the assessment of parameters in alkaline-polymer flooding using Bayesian models (Cheraghi et al., 2021; Zarin et al., 2025; Chicco et al., 2021), as well as the optimization of CO_2-EOR processes. The use of neural networks, ensemble methods, and hybrid algorithms has significantly improved the reliability of predictions, reduced the number of required laboratory tests, and led to the concept of intelligent digital laboratories.

Thus, machine learning is now considered a key element in modern experimental petroleum engineering, facilitating the transition from empirical approaches to predictive and adaptive modeling of core flooding laboratory processes.

Literature review. The analysis of relative permeabilities is one of the key tasks in laboratory studies of filtration processes in porous media. Classical approaches based on analytical models, such as the Corey model (Corey, 1954), are widely used for approximating experimental data. However, these methods assume a fixed functional form and require the adjustment of empirical coefficients, which limits their applicability in the presence of complex nonlinear relationships between system parameters.

Modern studies demonstrate a transition from classical empirical models to data-driven approaches. In particular, machine learning methods are actively used to predict filtration characteristics, including relative permeabilities and oil recovery factors (Alali et al., 2025; Zarin et al., 2025; Kalam et al., 2020; Talebkeikhah et al., 2021). It has been shown that the application of machine learning algorithms significantly improves approximation accuracy compared to traditional methods (Zhao et al., 2020; Reginato et al., 2023).

One of the most common approaches is the use of regression models. Linear and multivariable regression are applied to evaluate relationships between core properties and fluid characteristics. Despite their simplicity and interpretability, these models demonstrate limited accuracy in the presence of strong nonlinearities typical of two-phase flow processes (Ahmadi & Chen, 2019; Khilrani et al., 2021).

Therefore, particular attention has been given to ensemble methods such as Random Forest and AdaBoost. These algorithms are capable of capturing complex nonlinear dependencies and exhibit high robustness to noise in experimental data. According to several studies, ensemble models provide high coefficients of determination ($R^2 > 0.95$) when predicting relative permeabilities (Alali et al., 2025; Kalam et al., 2020; Yoga et al., 2024). In addition, they effectively handle high-dimensional datasets by accounting for the influence of multiple physical factors, including permeability, fluid viscosity, and capillary effects (Zhou et al., 2024; Al Khalifah et al., 2020).

Significant progress has also been achieved in hybrid approaches that combine machine learning with physical models. Such methods improve prediction stability and ensure physical interpretability by integrating experimental and numerical data (Zhao et al., 2020; Reginato et al., 2023; Mukherjee, 2025).

A separate research direction involves the application of machine learning to process computed tomography (CT) data. The integration of pore structure characteristics with filtration parameters enhances the accuracy of permeability and relative permeability predictions, particularly in heterogeneous reservoirs (Zarin et al., 2025; Hussen et al., 2024).

In addition, probabilistic and Bayesian approaches are actively developed for uncertainty quantification and parameter estimation in displacement models. These methods are particularly relevant for analyzing chemical enhanced oil recovery processes (Steineder et al., 2020; Cheraghi et al., 2021).

Overall, the analysis of global literature shows that machine learning is becoming a key tool in the interpretation of laboratory core flooding experiments. Its application enables capturing nonlinear relationships between system parameters, reducing the volume of laboratory studies, improving accuracy and reproducibility, and accelerating data processing (Wang & Chen, 2023; Zhou et al., 2024).

Despite significant progress, several challenges remain, including limited availability of experimental data, the risk of model overfitting, and the need to ensure physical interpretability of the results. These challenges define the relevance of further research aimed at developing robust and physically consistent machine learning approaches for laboratory filtration studies.

Materials and methods. Machine learning (ML) is a set of algorithms capable of identifying patterns in data and constructing mathematical models that can predict system behavior without explicitly defining physical-mathematical equations (Corey, 1954; Zhao et al., 2020).

In traditional programming, data and rules are provided to the computer, and the expected answer is calculated using these rules. However, in ML, both the data and the expected answers are provided to the model, and the rules are the result. These learned rules are then applied to predict the response for new data. The difference between ML and traditional programming is illustrated in Figure 1 (Cheraghi et al., 2021).

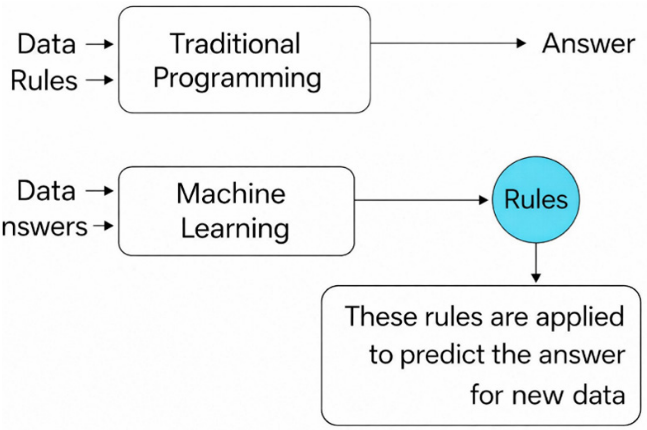

Figure 1. Difference Between Machine Learning and Traditional Programming.

In laboratory core flooding experiments, machine learning (ML) is used for processing and analyzing experimental data, predicting filtration characteristics, assessing parameter sensitivity, and optimizing injection conditions. This allows for complementing, rather than replacing, traditional physical experiments with intelligent analytical methods (Hussen et al., 2024).

Traditional analysis of laboratory core flooding data involves repeated approximation of relative permeability curves and model coefficient adjustments. Machine learning can replace this labor-intensive step by automatically identifying relationships between S_w, porosity, permeability, viscosities, and the k_{rw}/k_{ro} coefficients. Thus, ML does not replace the physics of the process but simplifies the mathematical part of the analysis, reducing processing time and improving result reproducibility.

A typical sequence of ML application in laboratory research includes:

1. Data collection and preprocessing (filtering, normalization, noise removal).

2. Model training based on core parameters, fluid properties, and injection conditions.

3. Accuracy assessment using statistical metrics (R^2, RMSE, MAE, MAPE).

4. Interpretation and optimization, i.e., determining the optimal combinations of pressure, concentration, and displacement rate.

Main machine learning methods used in core flooding laboratory experiments:

1. Regression Models. Linear and multivariable regression are widely used for predicting minimum miscibility pressure (MMP), relative permeabilities, and other filtration characteristics obtained in core flooding laboratory experiments. In such studies, regression models are used to refine the relationships between measured parameters such as pressure, saturation, and viscosity, as well as for approximating relative permeability curves.

In practice, regression analysis is often part of combined computational schemes (e.g., using ensemble models), which increases the consistency between predicted and experimental data and achieves high values of the coefficient of determination ($R^2 > 0.95$) (Kalam et al., 2020; Massarweh et al., 2026; Xue et al., 2025; Ahmadi & Chen, 2019; Khilrani et al., 2021; Reginato et al., 2023; Wang & Chen, 2023; Khosravi et al., 2024; Mahdy et al., 2024; Yoga et al., 2024).

2. Ensemble Methods. Ensemble methods, including Random Forest and AdaBoost algorithms, have shown high resistance to noise and nonlinear dependencies typical of laboratory data on oil displacement.

In several studies, applying ensemble approaches has allowed for reproducing experimental relative permeability curves and the dependency of the oil recovery factor on injection conditions with high accuracy.

These algorithms are particularly effective when analyzing large sets of laboratory data, where it is necessary to account for the influence of numerous physical factors, including permeability, fluid viscosity, and capillary effects (Alali et al., 2025; Xue et al., 2025; Zhou et al., 2024; Al Khalifah et al., 2020).

Table 1. Comparison of Machine Learning Methods for Core Flooding Experiments.

ML Method	Application in Laboratory Experiments	Advantages	Limitations	Typical Accuracy Metrics
Linear and Multivariable Regression	Prediction of MMP, relative permeabilities	Simplicity, high interpretability	Poor performance with nonlinear dependencies	$R^2 = 0.85–0.90$
Random Forest	Waterflooding and polymer flooding	Resistance to noise, ability to handle large numbers of features	Requires a large volume of data	$R^2 > 0.95$
AdaBoost (in ensemble)	Prediction of relative permeabilities	Increases accuracy of the final ensemble	Possible overfitting effect with small data	Error ≤ 5%

Results and Discussion. The relevance of this study lies in the need to improve the methods for analyzing and interpreting laboratory core flooding experiments to enhance the accuracy of evaluating filtration characteristics and the effectiveness of enhanced oil recovery technologies. Traditional approaches do not always account for nonlinear dependencies between physicochemical parameters, whereas the use of machine learning methods enables deeper data processing, system behavior prediction, and optimization of injection conditions. The aim of this work is to summarize global experience in applying machine learning in laboratory core flooding experiments, analyze the effectiveness of various algorithms, and determine the potential for their integration into laboratory practices to increase the reliability and reproducibility of results.

The results of this study will expand the understanding of machine learning capabilities in interpreting laboratory experiments and provide recommendations for its implementation in laboratories involved in core flooding process modeling. This will ensure more accurate estimation of oil displacement parameters, speed up result analysis, and reduce the costs of conducting laboratory series.

One of the key directions for the implementation of ML is the prediction of phase relative permeabilities and oil recovery factors based on laboratory measurements.

In most studies, statistical accuracy metrics were used to assess the quality of the models, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE):

1. The Coefficient of Determination (R^2) is defined by the formula (1):

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(X_i - Y_i)^2}{\sum_{i=1}^{m}(\bar{Y}_i - Y_i)^2} \quad (1)$$

where: R^2 – coefficient of determination (R-squared);
X_i – predicted value (prediction);
Y_i – actual value (ground truth);
$\bar{Y}$ – mean of the actual values;
m – number of observations.
The best value is = +1, and the worst value is = −∞ (Kalam et al., 2020).

2. The Root Mean Square Error (RMSE) is defined by the formula (2):

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \quad (2)$$

where: $RMSE$ – Root Mean Square Error (RMSE);
X_i – predicted value (prediction);
Y_i – actual value (ground truth);
m – number of observations.
The best value is = 0, and the worst value is = +∞ (Kalam et al., 2020).

3. The Mean Absolute Percentage Error (MAPE) is defined by the formula (3):

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{Y_i - X_i}{Y_i}\right| \qquad (3)$$

where: $MAPE$ – Mean Absolute Percentage Error (MAPE);
Y_i – actual value (ground truth);
X_i – predicted value (prediction);
m – number of observations.
The best value is = 0, and the worst value is = +∞ (Kalam et al., 2020).

The use of these metrics not only quantitatively confirmed the accuracy of predictions but also revealed the patterns that determine the influence of input parameters on filtration results (Sadrikhanloo et al., 2026; Tran et al., 2020; Matinkia et al., 2023; Erofeev et al., 2019).

Figure 2 presents a comprehensive analysis of the relative importance of input parameters when predicting water relative permeability (Krw) and oil relative permeability (Kro). These dependencies help identify which physicochemical properties of the rock and fluids have the greatest influence on the prediction outcome. This is particularly important when interpreting laboratory data with high uncertainty and variability (Hussen et al., 2024).

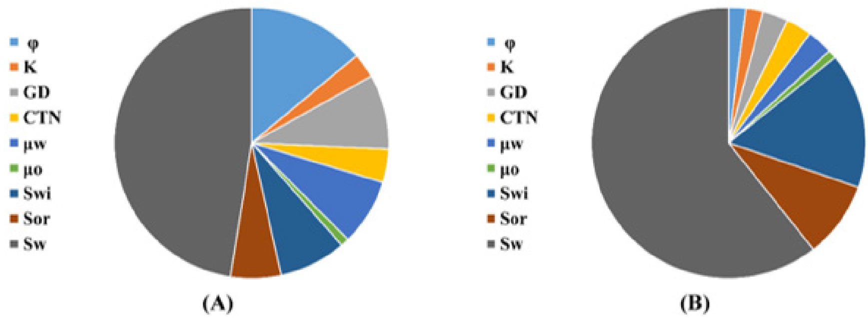

Figure 2. Parameter Importance in Predicting Relative Permeability (A) k_{rw} of Water and (B) k_{ro} of Oil.

Diagram (A) shows that when predicting water relative permeability (k_{rw}), the most significant factor is water saturation (S_w), contributing to more than 40% of the total importance. This aligns with the physical nature of the process: as S_w increases, oil is intensively displaced from the pore space, directly affecting the filtration capacity of the water phase.

Additionally, notable contributions to the prediction accuracy are made by porosity (φ) and permeability (k), which determine the structure and connectivity of the core's pore space. The average value of computed tomography scans (CTN) and grain density (GD) play a secondary role, influencing the distribution of saturations and the shape of the filtration curves.

Diagram (B), which represents the relative permeability of oil (k_{ro}), also shows that water saturation (S_w) is the dominant parameter, accounting for more than 50% of the importance. Significant influence is exerted by residual oil saturation (S_{or}) and oil viscosity (μ_o),, which are related to flow resistance and changes in the flow regime of oil in the capillary channels.

Porosity (φ) and initial water saturation (S_{wi}) parameters have a lesser effect on the shape of the Kro curves but contribute to stabilizing the model's accuracy, especially at early stages of flooding.

Overall, the results of the feature importance analysis confirm that the most significant factors in predicting filtration characteristics are water saturation, fluid viscosity, and rock permeability. These parameters reflect both the physical foundation of oil displacement processes and the sensitivity of the machine learning model to changes in phase properties.

Thus, the constructed diagrams visually demonstrate that ML models not only replicate experimental dependencies but also allow for the quantitative assessment of each parameter's contribution, thereby enhancing the physical interpretability of the obtained results.

Demonstration Numerical Experiment on the Application of ML to Core Data.

To assess the applicability of machine learning (ML) methods for core data analysis, a controlled numerical experiment was conducted in the Python environment using specialized libraries for data processing and regression model construction (Pedregosa et al., 2011). The task was to verify the ability of the ML model to reproduce nonlinear dependencies of relative permeabilities on rock parameters and filtration conditions without explicitly using analytical equations for two-phase flow (Mukherjee, 2025).

Since access to full laboratory data may be limited, a synthetic dataset was generated within realistic parameter ranges. This approach allowed testing the ML methodology under a fully controlled physical basis and eliminated uncertainties related to the quality of experimental measurements.

A dataset of 1200 rows with 7 parameters was formed. Each observation corresponds to a conditional core measurement under given rock parameters and filtration conditions.

The following initial parameters were used:

– φ – porosity (0.12–0.32);

– k_{abs} – absolute permeability (lognormal distribution, with an average permeability of 150 mD and variability);

– S_w – water saturation (0.15–0.85);

– μ_o – в oil viscosity, cP (2–20);

– μ_w – water viscosity, cP (0.3–1.5);

– ΔP – pressure drop, MPa (0.2–3.0, normalized units);

– CTN – heterogeneity index (0–1).

The parameter ranges correspond to typical conditions in laboratory core research.

The target variables were:

– k_{rw}– water relative permeability;

– k_{ro} – oil relative permeability.

To generate these, a modified Corey-type model was used (Corey, 1954), according to formulas (4, 5):

$$k_{rw} = k_{rw0} \cdot S_e^{n_w} \tag{4}$$

$$k_{ro} = k_{ro0} \cdot (1 - S_e)^{n_o} \tag{5}$$

Where the effective saturation is defined by formula (6):

$$S_e = \frac{S_w - S_{wc}}{1 - S_{wc} - S_{or}} \tag{6}$$

The values of residual saturations were taken as: $S_{wc} = 0.20$ and $S_{or} = 0.25$, which correspond to typical average laboratory characteristics.

To increase realism, the parameters of the power indices n_w, n_o, as well as the critical values k_{rw0}, k_{ro0}, were weakly correlated with porosity and the heterogeneity index. Additionally, a small random noise (~3%) was added to the calculated relative permeability values to simulate experimental errors.

It should be emphasized that the Corey analytical model was used exclusively for generating reference values. The machine learning model did not have access to the analytical equations.

The dataset was split into a training set (75%) and a testing set (25%). Two separate regression models were built:

– Model A – prediction of k_{rw}.

– Model B – prediction of k_{ro}.

The algorithm used was the Random Forest ensemble method, which has the ability to approximate complex nonlinear dependencies without the need to specify a functional form.

The model was trained based on the following data format:

$$(\varphi,\ k_{abs},\ S_w,\ \mu_o,\ \mu_w,\ \Delta P,\ CTN) \rightarrow k_r$$

The model quality was assessed using the following metrics:

– Coefficient of determination (R^2).

– Root Mean Square Error (RMSE).

The ML model used only input parameters and was trained based on "input → output" examples.

Figure 3 presents a comparison of predicted (Y-axis) and reference (X-axis) values of k_{rw}, for the test set. The concentration of points along the 1:1 diagonal indicates high accuracy in reproducing the dependency. A similar result was obtained for k_{ro}, as shown in Figure 4.

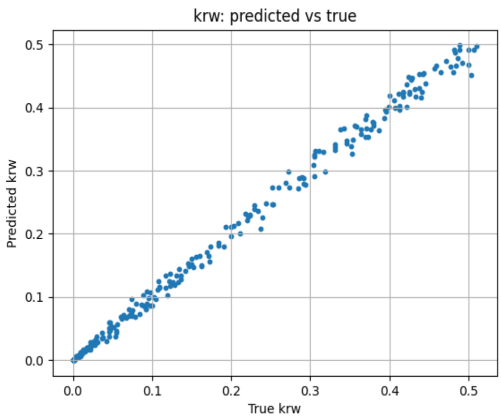

Figure 3. Comparison of predicted and reference values of k_{rw} for the test set.

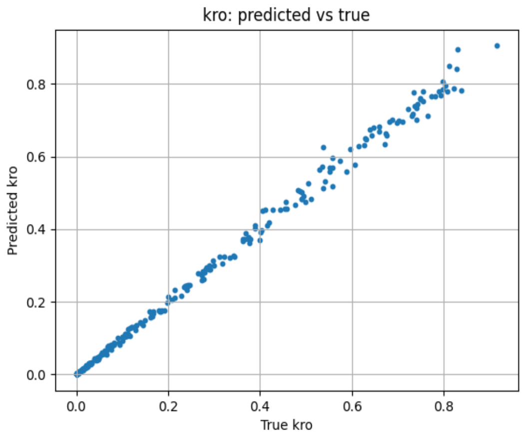

Figure 4. Comparison of predicted and reference values of k_{ro} for the test set.

The obtained values of the coefficient of determination are close to one, which confirms the ensemble model's ability to accurately approximate the nonlinear dependencies embedded in the data.

Figures 5 and 6 present the comparison of $k_r(S_w)$ dependencies for selected synthetic cores. The machine learning model correctly reproduces the shape of the relative permeability curves, including their pronounced nonlinearity, capturing their characteristic dependency.

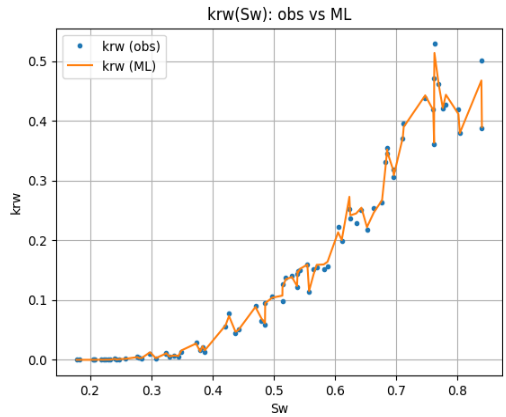

Figure 5. Reproduction of the ML model's water relative permeability curves as a function of water saturation $k_{rw}(S_w)$.

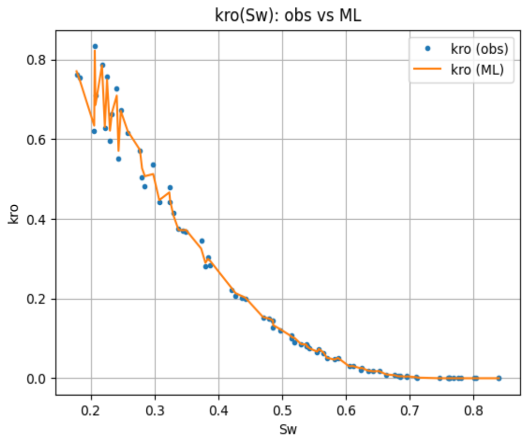

Figure 6. Reproduction of the ML model's oil relative permeability curves as a function of water saturation $k_{ro}(S_w)$.

This result is fundamentally important, as it confirms not only the numerical agreement of values but also the correct reproduction of the physical structure of the dependency on saturation.

Figure 7 presents the feature importance evaluation (Permutation importance). The greatest contribution to the prediction is made by water saturation (Sw), followed by porosity (φ) and the heterogeneity index (CTN). While the Corey model relies solely on saturation and empirical power indices, the Random Forest model independently identified that rock structural characteristics carry additional predictive value, capturing broader physical relationships present in the data. Thus, ML not only predicts but also confirms the physical interpretation of two-phase filtration, demonstrating the physical interpretability and consistency of the constructed model.

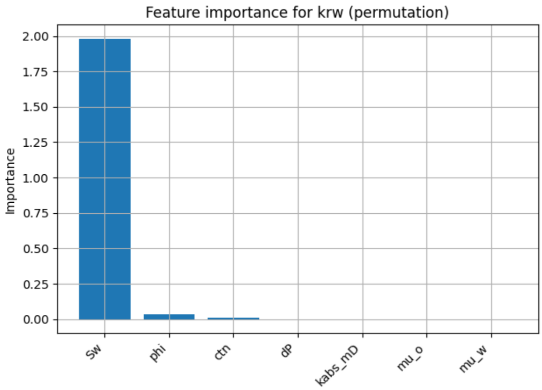

Figure 7. Feature importance evaluation (Permutation importance) for the ML model.

The obtained feature importance structure aligns with the physics of two-phase filtration, where relative permeability is primarily determined by saturation and the structure of the pore space. Thus, the ML model demonstrates interpretability and physical consistency.

In the conducted numerical experiment, the coefficient of determination was $R^2 = 0.996$ for k_{rw} and $R^2 = 0.994$ for k_{ro}. The root mean square error (RMSE) did not exceed 0.012 for both variables, which is significantly lower than the range of relative permeability values (0–1). The results indicate high accuracy in approximating the nonlinear dependency.

The MAPE metric was not used for model quality analysis because relative permeabilities can approach zero. In such cases, relative error becomes disproportionately high and does not reflect the true approximation accuracy. For this task, R^2 and RMSE metrics are more informative.

The demonstration experiment confirms the feasibility of using machine learning methods to recover complex nonlinear dependencies between rock parameters and relative permeabilities.

The developed numerical experiment demonstrates the technical feasibility of integrating ML into core data analysis. When using real core flooding data, this approach can be applied to:

– interpretation of experimental data;
– sensitivity analysis of parameters;
– reduction in the volume of laboratory studies;
– construction of digital analogs of core experiments;
– accelerated prediction of relative permeabilities;
– identification of dominant filtration factors;
– preparation of input data for hydrostatic modeling.

The results confirm the promising use of machine learning methods in core research analysis and interpretation tasks.

Conclusions. Machine learning has proven to be an effective tool for analyzing and predicting the results of core flooding laboratory experiments, improving accuracy and reducing uncertainty in the interpretation of filtration characteristics.

1. The application of regression algorithms, ensemble methods, and probabilistic models allowed for the reproduction of experimental dependencies between the parameters of porous media and fluids with a high coefficient of determination ($R^2 > 0.9$).

2. Feature importance analysis showed that key factors influencing relative permeability of oil and water include water saturation, fluid viscosity, and rock permeability, confirming the physical validity of the ML models.

3. The use of R^2, RMSE, and MAPE metrics enabled an objective assessment of prediction quality and highlighted the advantages of ensemble models (Random Forest, AdaBoost) when working with nonlinear and noisy experimental data.

4. The integration of machine learning into laboratory practice contributes to reducing the volume of experimental trials, automating data processing, and creating digital twins of laboratory processes.

5. Global experience demonstrates that the integration of ML in core flooding research opens up prospects for creating next-generation intelligent laboratories, where physical experiments and computational modeling complement each other.

References

Agzamov A., Efendiyev G., Moldabayeva G.Zh., Syzdykov A., Suleimenova R., Tuzelbayeva Sh., Zaurbekov K. (2022) On the degree of influence of waterflooding on the oil recovery factor from productive formations of high-viscosity reservoirs X, represented by terrigenous reservoirs. – Kompleksnoe Ispolzovanie Mineralnogo Syra = Complex Use of Mineral Resources. – Vol. 326. – No. 3. – P. 50-58. https://doi.org/10.31643/2023/6445.28 (in Eng.).

Ahmadi M.A., Chen Z. (2019) Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. – Petroleum. – Vol. 5. – No. 3. – P. 271–284. https://doi.org/10.1016/j.petlm.2018.06.002 (in Eng.).

Al Khalifah H., Glover P.W.J., Lorinczi P. (2020) Permeability prediction and diagenesis in tight carbonates using machine learning techniques. – Marine and Petroleum Geology. – Vol. 112. – Art. 104096. https://doi.org/10.1016/j.marpetgeo.2019.104096 (in Eng.).

Alali R., Abe K., Fujii H. (2025) Simulation of Core Flooding with Predicted Oil and Water Relative Permeabilities Using Bagging, Boosting, and Stacking Machine Learning Techniques. – ACS Omega. – Vol. 10. – No. 16. – P. 15967–15978. https://doi.org/10.1021/acsomega.4c07089 (in Eng.).

Cheraghi Y., Kord S., Mashayekhizadeh V. (2021) Application of machine learning techniques for selecting the most suitable enhanced oil recovery method; challenges and opportunities. – Journal of Petroleum Science and Engineering. – Vol. 205. – Art. 108761. https://doi.org/10.1016/j.petrol.2021.108761 (in Eng.).

Chicco D., Warrens M.J., Jurman G. (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. – PeerJ Computer Science. – Vol. 7. – Art. e623. DOI: 10.7717/peerj-cs.623 (in Eng.).

Corey A.T. (1954) The interrelation between gas and oil relative permeabilities. – Producers Monthly. – Vol. 19. – No. 1. – P. 38–41. (in Eng.).

Erofeev A., Orlov D., Ryzhov A., Koroteev D. (2019) Prediction of Porosity and Permeability Alteration Based on Machine Learning Algorithms. – Transport in Porous Media. – Vol. 128. – No. 2. – P. 677–700. DOI: 10.1007/s11242-019-01265-3 (in Eng.).

Ginsburg B. (2017) Application case study – machine learning. In: Kirk D.B., Hwu W.W.

Programming Massively Parallel Processors. – Elsevier. – P. 345–367. DOI: 10.1016/B978-0-12-811986-0.00016-9 (in Eng.).

Hussen A., Munshi T.A., Jahan L.N., Hashan M. (2024) Advanced machine learning approaches for predicting permeability in reservoir pay zones based on core analyses. – Heliyon. – Vol. 10. – No. 12. – Art. e32666. https://doi.org/10.1016/j.heliyon.2024.e32666 (in Eng.).

Kalam S., Khan M., Khan R.A., Alam M.M.M., Sadeed A., Mahmoud M., Abu-Khamsin S.A. (2020) Advancing Relative Permeability Estimation Through Data-Driven Modeling. – SPE. https://doi.org/10.2118/202810-MS (in Eng.).

Khilrani N., Prajapati P., Patidar A.K. (2021) Contrasting machine learning regression algorithms used for the estimation of permeability from well log data. – Arabian Journal of Geosciences. – Vol. 14. – No. 20. – Art. 2070. DOI: 10.1007/s12517-021-08390-8 (in Eng.).

Khosravi R., Simjoo M., Chahardowli M. (2024) Low salinity water flooding: estimating relative permeability and capillary pressure using coupling of particle swarm optimization and machine learning technique. – Scientific Reports. – Vol. 14. – No. 1. – Art. 13213. DOI: 10.1038/s41598-024-61168-2 (in Eng.).

Mahdy A., Zakaria W., Helmi A., Helaly A.S., Mahmoud A.M.E. (2024) Machine learning approach for core permeability prediction from well logs in Sandstone Reservoir, Mediterranean Sea, Egypt. – Journal of Applied Geophysics. – Vol. 220. – Art. 105249. https://doi.org/10.1016/j.jappgeo.2023.105249 (in Eng.).

Massarweh O., Abd A.S., Schneider J., Abushaikha A.S. (2026) Application of machine learning for permeability prediction in heterogeneous carbonate reservoirs. – Artificial Intelligence in Geosciences. – Vol. 7. – No. 1. – Art. 100183. https://doi.org/10.1016/j.aiig.2025.100183 (in Eng.).

Matinkia M., Hashami R., Mehrad M., Hajsadeedi M.R., Velayati A. (2023) Prediction of permeability from well logs using a new hybrid machine learning algorithm. – Petroleum. – Vol. 9. – No. 1. – P. 108–123. https://doi.org/10.1016/j.petlm.2022.03.003 (in Eng.).

Mukherjee S. (2025) Relative Permeability Modeling: A Synergy of Machine Learning, Pore Network Modeling, and Field Data Analysis. PhD dissertation. The Pennsylvania State University, The J. Jeffrey and Ann Marie Fox Graduate School. (in Eng.).

Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., Duchesnay E. (2011) Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. – Vol. 12. – P. 2825–2830. (in Eng.).

Reginato L.F., Gioria R.S., Sampaio M.A. (2023) Hybrid Machine Learning for Modeling the Relative Permeability Changes in Carbonate Reservoirs under Engineered Water Injection. – Energies. – Vol. 16. – No. 13. – Art. 4849. https://doi.org/10.3390/en16134849 (in Eng.).

Sadrikhanloo S., Busch B., Hilgers C. (2026) Porosity and permeability prediction from petrographic point-counting data using machine learning: Applications to Rotliegendes and Buntsandstein reservoirs. – Energy Geoscience. – Vol. 7. – No. 2. – Art. 100537. https://doi.org/10.1016/j.engeos.2026.100537 (in Eng.).

Steineder D., Vanegas G., Clemens T., Zechner M. (2020) Deriving Alkali Polymer Parameter Distributions from Core Flooding by Applying Machine Learning in a Bayesian Framework to Simulate Incremental Oil Recovery. – SPE. https://doi.org/10.2118/200578-MS (in Eng.).

Talebkeikhah M., Sadeghtabaghi Z., Shabani M. (2021) A Comparison of Machine Learning Approaches for Prediction of Permeability using Well Log Data in the Hydrocarbon Reservoirs. – Journal of Human, Earth, and Future. – Vol. 2. – No. 2. – P. 82–99. https://doi.org/10.28991/HEF-2021-02-02-01 (in Eng.).

Tran H., Kasha A., Sakhaee-Pour A., Hussein I. (2020) Predicting carbonate formation permeability using machine learning. – Journal of Petroleum Science and Engineering. – Vol. 195. – Art. 107581. https://doi.org/10.1016/j.petrol.2020.107581 (in Eng.).

Wang H., Chen S. (2023) Insights into the Application of Machine Learning in Reservoir Engineering: Current Developments and Future Trends. – Energies. – Vol. 16. – No. 3. – Art. 1392. https://doi.org/10.3390/en16031392 (in Eng.).

Xue W., Chen Z., Jia C., Wang H. (2025) Big data-driven optimization of combined water flooding and hydraulic fracturing in oilfields: Synergistic effect analysis and recovery enhancement strategies. – Advances in Resources Research. – Vol. 5. – No. 3. – P. 1217–1239. DOI: 10.50908/arr.5.3_1217 (in Eng.).

Yoga H.F., Johns R.T., Purswani P. (2024) Predictive Model for Relative Permeability Using Physically-Constrained Artificial Neural Networks. – SPE Journal. – Vol. 29. – No. 2. – P. 928–942. DOI: 10.2118/209420-PA (in Eng.).

Zarin T., Ahmadi Eshkaftaki H., Sharifi A. (2025) Machine learning-based prediction of oil–water relative permeability using core flooding and CT-scan data. – Results in Engineering. – Vol. 27. – Art. 105735. https://doi.org/10.1016/j.rineng.2025.105735 (in Eng.).

Zhao B., Ratnakar R., Dindoruk B., Mohanty K. (2020) A Hybrid Approach for the Prediction of Relative Permeability Using Machine Learning of Experimental and Numerical Proxy SCAL Data. – SPE Journal. – Vol. 25. – No. 5. – P. 2749–2764. https://doi.org/10.2118/196022-PA (in Eng.).

Zhou W., Liu C., Liu Y., Zhang Z., Chen P., Jiang L. (2024) Machine Learning in Reservoir Engineering: A Review. – Processes. – Vol. 12. – No. 6. – Art. 1219. https://doi.org/10.3390/pr12061219 (in Eng.).

Publication Ethics and Publication Malpractice in the journals of the Central Asian Academic Research Center LLP

For information on Ethics in publishing and Ethical guidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the journals of the Central Asian Academic Research Center LLP implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier. com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The Central Asian Academic Research Center LLP follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics. org/files/ u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/ or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP will monitor and safeguard publishing ethics.

Requirements for articles design for publication in the journal are available on the websites:

www:nauka-nanrk.kz
http://www.geolog-technical.kz/index.php/en/
ISSN 2518-170X (Online),
ISSN 2224-5278 (Print)

Managing Editor: *T. Apendiev*
Editors: *D.S. Alenov, A.Shormakova*
Computer layout: *G.D. Zhadyranova*

Signed for print: July 10, 2026
Format: 70×90 1/16. 26.5 printed sheets. Order No. 3.

«Central Asian Academic Research Center» LLP
Almaty, Shevchenko Street, 28