Energy Geoscience 7 (2026) 100537

Contents lists available at ScienceDirect

Energy Geoscience

journal homepage: www.keaipublishing.com/en/journals/energy-geoscience

Porosity and permeability prediction from petrographic point-counting data using machine learning: Applications to Rotliegendes and Buntsandstein reservoirs

Sahar Sadrikhanloo, Benjamin Busch *, Christoph Hilgers

Structural Geology and Tectonics, Institute of Applied Geosciences, Karlsruhe Institute of Technology, Adenauerring 20a, 76131, Karlsruhe, Germany

A R T I C L E I N F O

Article history:
Received 5 August 2025
Received in revised form
4 December 2025
Accepted 30 January 2026

Keywords:
Machine learning
Petrography
Point-counting
Porosity
Permeability

A B S T R A C T

Machine learning approaches are widely used in geosciences. However, one widely available dataset in reservoir geology remains underrepresented in published works: petrographic data from classical point-counting analyses. Such data are widely available for reservoir lithology characterization, often in combination with routine core analysis data (porosity and permeability). Since porosity and permeability in siliciclastic rocks are controlled by the detrital and authigenic composition and samples record effects of compaction during diagenesis, these datasets are often linked to assess reservoir quality controls.

Datasets from six wells, covering four regions and two large reservoir lithologies in central Europe, the Permian Rotliegendes and Triassic Buntsandstein, were used to apply machine learning to the petrographic and reservoir quality data to predict porosity and permeability. Predictions are based on point-counting data including detrital and authigenic phases, optical porosity, grain-to-IGV (GTI) and grain-to-grain (GTG) coating coverages, and granulometry. For both regression tasks, a Random Forest and a Support Vector Regression machine learning model were implemented, with performance compared and the best model selected based on coefficient of determination (R^2) and error metrics. Porosity predictions using a Random Forest algorithm yielded an R^2 of 0.92, a mean average error (*MAE*) of 1.25%, and a root mean square error (*RMSE*) of 1.56%. Permeability predictions of real-scale permeability using Support Vector Regression gave an R^2 of 0.85, *MAE* of 29.4 mD, *RMSE* of 68.3 mD, and a range-based normalized *RMSE* of 8.76% (real-scale). Log-transformation of measured and predicted permeability resulted in a more representative R^2 of 0.83, *MAE* of 0.21, and *RMSE* of 0.24, reflecting its log-normal distribution. Predictions are acceptable despite the limited dataset, which reduces operator bias by using curated data. This machine learning approach may simultaneously unlock another understanding of reservoir quality controls based on SHapley Additive exPlanations (SHAP) value plots.

Further training of such models on cored reservoir sections can improve understanding of which detrital and authigenic mineral phases influence reservoir properties. Trained models could also potentially evaluate reservoir properties from cuttings, which, like well logs, are more continuous than cores while allowing diagenetic interpretation based on petrographic analysis.

1. Introduction

Machine learning approaches have widely been applied to the geological subsurface in the past years (Bergen et al., 2019; Karpatne et al., 2019; Dramsch, 2020). Their applications include well log analysis and derivation of physical rock properties (Mishra et al., 2022; Ore and Gao, 2023), pore classification (Yang et al., 2025), porosity and permeability assessment during drilling (Sun et al., 2021), seismic interpretation (Wrona et al., 2018), and facies classification (Dubois et al., 2007; Duarte et al., 2023). Some reservoir-relevant data, such as porosity, permeability, or lithology/facies, can be derived from well logs using machine learning and are already applied in reservoir geology (Wood, 2020; Rahimi

* Corresponding author.
E-mail address: benjamin.busch@kit.edu (B. Busch).
Peer review under the responsibility of Editorial Board of Energy Geoscience.

https://doi.org/10.1016/j.engeos.2026.100537

and Riahi, 2022; Zanganeh Kamali et al., 2022; Abbas et al., 2023). However, underlying diagenetic factors relating to the development of reservoir quality differences are not well addressed. This is largely related to the input data being geophysical logs, which can indicate some (but not all) mineral phases present in a lithology. Therefore, an understanding of diagenetic reservoir quality controls cannot be accurately addressed and still relies on the integration of petrographic and petrophysical data. Simultaneously, machine learning approaches are applied to assess petrographic sample suites and derive the mineralogical composition, porosity distribution, capillary pressure curves, or their relation to petrofacies (Rubo et al., 2019; Tang et al., 2020; Saxena et al., 2021; Johnson et al., 2022; Duarte et al., 2023; Yu et al., 2023; Esmaeili et al., 2024). However, petrographic data is already widely available for reservoir lithologies around the globe and can be linked to available porosity and permeability datasets. While well logs cannot completely capture detrital and authigenic minerals, and current petrographic image-based machine learning approaches do not yet distinguish the effect of different mineral textures and the presence of certain minerals at grain interfaces, petrographic point-counting data can bridge that gap. Leveraging this resource through machine learning (as a data science application) could reduce drilling costs while maintaining insight into reservoir quality and its controls. Using machine learning approaches and applying them to available legacy datasets may be a suitable tool to extend their applicability to other aspects of reservoir geology, especially in light of the energy transition and an extended geothermal or storage utilization of fractured and porous reservoir rocks.

Standard reservoir quality analyses still rely mostly on petrophysical (routine and special) core analysis data in conjunction with macroscopic and microscopic analyses, as well as correlation to well log measurements (Ajdukiewicz and Lander, 2010; Rider and Kennedy, 2011; Worden et al., 2018). Gained information on depositional environments and observed mineral distributions, textures, and their paragenetic relations aid in assessing controlling factors on reservoir quality.

In this study, we develop a workflow that integrates quantitative petrographic point count data with supervised machine learning to predict porosity and permeability. The workflow consists of assembling standardized petrographic and petrophysical datasets, defining the petrographic input features, and comparing two regression models: a Random Forest model and a Support Vector Regression model for porosity and permeability prediction. The models are optimized through hyperparameter tuning and evaluated using k-fold cross validation to ensure robust performance despite the limited dataset size. The two best-performing models are then selected to discuss model performance and geological controls. Therefore, SHapley Additive exPlanations (SHAP) values are used to interpret how individual petrographic attributes influence the predictions (Lundberg and Lee, 2017). If these effects on model predictions are combined with geological reasoning, they may uncover yet unknown reservoir quality controls.

If a suitable calibration can be achieved, such models could also be used to gain porosity and permeability estimates from petrographic samples prepared from e.g., cutting material obtained from lithologies on which the model is trained. Leveraging this untapped potential via machine-learning based data science may in future enable lower-cost drilling, while gaining comparable insights into reservoir quality and related controlling factors. This data can also be used in combination with well log data, which is mostly available along extended well sections, to optimally utilize all available data and material provided during drilling. This approach may be applied to reservoir systems across the world, if suitable calibration data is available.

2. Materials and methods

2.1. Geological input data

To test the applicability of machine learning on petrographic and petrophysical reservoir quality data, we selected a dataset covering core material from six wells (four wells from three regions targeting the Triassic Buntsandstein in southern Germany, two wells from one region targeting the Permian Rotliegendes in Northern Germany) (inset in Fig. 1a) (Monsees et al., 2020; Busch et al., 2022, 2024, 2025). The four sampled regions represent the western part of the Upper Rhine Graben [URG, Well A + B (Buntsandstein)], the eastern part of the URG [Well C (Buntsandstein)], the eastern shoulder of the URG [Well D (Buntsandstein)], and the Southern Permian Basin (Well A + B (Rotliegendes)] (inset in Fig. 1a). The overall dataset comprises 157 samples, covering a porosity range of 0–20% and a permeability range of 0.0001 mD to 780 mD (Fig. 1b). The porosity distribution across the four regions shows that well A + B (Buntsandstein) have the highest porosities, with one identified outlier, while well C exhibits the lowest porosity values (Fig. 1b). Well C (Buntsandstein) did not provide any permeability data, therefore the range of porosity values from that sample series is just highlighted in Fig. 1b. All utilized data is provided in supplementary materials 1.

The sandstone classification according to Folk (1980) reveals that samples from all studied wells fall into the subarkose, sublitharenite, lithic arkose, feldspathic litharenite, or litharenite classes (Fig. 1a). We selected these datasets, as they were analyzed by the same two petrographers maintaining comparability by utilizing the same comparator charts during data acquisition. They are also the only datasets continuously assessing grain coating coverages by different illite morphologies (radial and tangential) at different grain interfaces, which have been shown to affect reservoir properties (Heald, 1955; Heald and Larese, 1974; Pittman et al., 1992). Furthermore, they cover a wide range of present-day burial depths (400–4720 m TVD), to assess the model performance in relation to different present-day burial depths. However, as the degree of compaction and cementation recorded in the sandstones is a function of the burial history, and some samples were relatively uplifted to their paleo maximum burial depths, the present-day burial depth does not correlate to specific reservoir qualities in the studied samples. This is especially relevant for samples with the shallowest present-day burial depths [well D (Buntsandstein)], which have among the lowest average porosity and permeability of the studied samples. Furthermore, the effect of varying vertical effective stress histories (as recorded by the IGV) and thermal exposure and associated fluid-rock interaction (as recorded by contents of e.g., authigenic quartz and illite) are recorded in the individual rock microstructure assessed during point-counting to prepare the original input data. To test the robustness of the model we also use samples from two lithological groups from different arid fluvio-eolian depositional systems having different provenance areas during the Triassic and Permian, respectively.

The dataset comprises petrographic data (detrital and authigenic phases, optical porosity, GTI and GTG coating coverages, granulometry data) and petrophysical data (porosity and permeability) (supplementary materials I).

2.2. Machine learning setup

To assess whether porosity or permeability data are required to predict either of the other properties, they were included in the model setup, and their effect on model outcome is discussed.

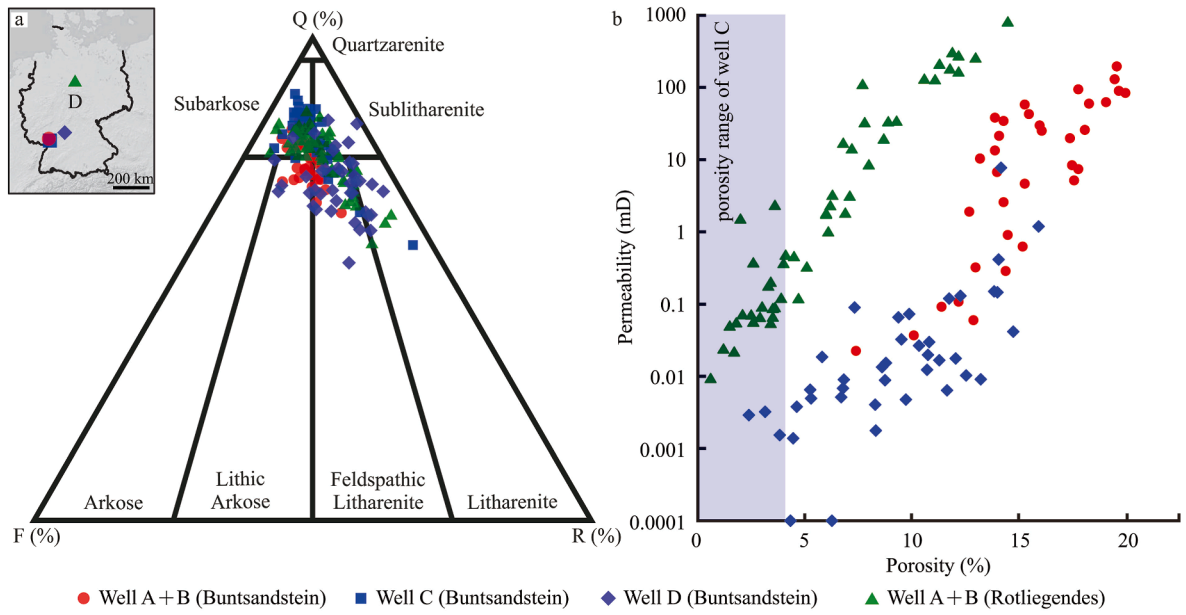

Fig. 1. QFR plot of the sample suite and map inset showing the distribution of the three studied regions (a) and porosity-permeability plot of the studied samples (b). Note that for well C (Buntsandstein), only porosity ranges are available and highlighted.

Further information on sample material and data can be gained from the original source publications. In total, the dataset contains 62 features with porosity and permeability designated as the target variables for the models.

At first, the dataset was randomly split into a training-validation set (80%) and a test set (20%) using sample wise shuffling. Due to the limited number of the datapoints and wells, well wise splitting was not enforced in this study. We employed Random Forest (RF) and Support Vector Regression (SVR) models to predict porosity and permeability and chose the best-performing for a more detailed assessment.

The employed RF model (Breiman, 2001) is an ensemble-based method that constructs multiple decision trees and averages their outputs (Cracknell and Reading, 2013; Bergen et al., 2019). RF is well suited for this study because it captures nonlinear relationships, interactions between features, and noise in high dimensional datasets. Hyperparameters including the number of estimators, maximum tree depth, minimum samples required to split a node, and minimum samples per leaf were optimized using Bayesian hyperparameter optimization (Brochu et al., 2010), evaluated through 5-fold cross validation. Cross validation performance was assessed using the coefficient of determination (R^2) as the objective metric.

SVR is a variant of support vector machines (Bergen et al., 2019) which identifies a hyperplane that maximizes the margin between predictions and an epsilon-insensitive tube (Cracknell and Reading, 2013; Shahnas et al., 2018). SVR is chosen as it can effectively captures relationships in small sample sets, where the number of features may exceed the number of samples (scikit-learn User Guide, 2025). One of the key drawbacks of Support Vector Machine algorithms is the increase in training time with an increase in samples (Guido et al., 2024). While this may not drastically affect this proof-of-concept study with a limited number of samples, this may affect larger sample suites. Hyperparameters including the regularization parameter (c), the insensitive loss width (ε), the kernel type (linear vs. radial basis function), and the kernel scale parameter (γ) were optimized using Bayesian hyperparameter optimization with 3-fold cross validation. For this dataset, which includes 123 permeability samples spanning several orders of magnitude, a linear kernel consistently outperformed in both cross validation and test evaluation.

Given the large range of permeability values across the dataset, a log-transformation was applied to the target variable before training the permeability model. This transformation helped stabilize variance and improve model interpretability. Additionally, RobustScaler was used to scale both the features and the log-transformed permeability values, ensuring robustness against outliers. The model performance was assessed using R^2, mean absolute error (*MAE*), and root mean square error (*RMSE*), as also suggested by Male and Duncan (2020). To facilitate a more interpretable assessment of prediction error on the original (real-scale) permeability values, we also report range-based normalized *RMSE* (*NRMSE*), which expresses the *RMSE* relative to the range of observed permeability values.

To evaluate whether feature reduction could improve generalization, we applied Recursive Feature Elimination (RFE) (Guyon et al., 2002) using the tuned models. RFE analyses demonstrated that cross validation and test set performance increased by adding the selected features up to 45 and then the change was not significant. To improve geological interpretability, all 60 petrographic features were retained.

Feature importance was analyzed using SHAP values, providing insights into the effect of specific properties on model predictions. The SHAP plots are based on the training data, as this dataset contains more samples and correlations are more clearly visible.

3. Results and discussion

3.1. Model selection

Based on the performance of machine learning models, we selected the best-performing model to present a more in-depth assessment of controlling factors in the following sections.

For the porosity prediction model, the RF algorithm slightly outperforms SVR in R^2, *MAE*, and *RMSE* (Table 1). While this

Table 1
Comparison of model metrics of the testing and training/validation datasets of the random forest and support vector models, including the standard deviation for the CV results. Correlation of the true and predicted log-transformed permeability values allows a more accurate assessment of errors for testing and cross-validation.

		Model					
		RF			SVR		
		Porosity (%)	Permeability (mD)	Log-transformed permeability	Porosity (%)	Permeability (mD)	Log-transformed permeability
Test	R^2	0.92	0.14	0.81	0.88	0.81	0.83
	MAE	1.25	48.97	0.20	1.50	29.44	0.21
	RMSE	1.56	146.13	0.25	1.90	68.37	0.24
Mean CV results	R^2	0.86 ± 0.06	0.32 ± 0.30	0.81 ± 0.07	0.84 ± 0.04	0.50 ± 0.38	0.77 ± 0.05
	MAE	1.46 ± 0.26	20.17 ± 10.56	0.19 ± 0.06	1.75 ± 0.14	18.33 ± 6.80	0.21 ± 0.02
	RMSE	1.93 ± 0.30	70.37 ± 25.41	0.24 ± 0.07	2.14 ± 0.15	42.81 ± 16.90	0.24 ± 0.02

Notes: CV denotes cross-validation, RF denotes Random Forest, SVR denotes Support Vector Regression.

indicates that both algorithms may suitably predict porosity, with acceptable errors, the Random Forest model is selected to be presented in more detail in section 3.2. However, for the permeability prediction in real-space (i.e. using log-normally distributed permeability values), the SVR algorithm outperforms the RF algorithm (Table 1). Log-transforming the true and predicted permeability from the test dataset results in more comparable, but also more reasonable error metrics with *MAE* and *RMSE*, which are very similar for both models. Due to its higher R^2, the SVR model was selected for subsequent presentation in section 3.4.

Comparing the mean results from cross-validation with the testing metrics, a similar spread of error metrics for training/validation and test datasets can be observed for porosity and log-transformed permeability prediction. Due to the log-normal distribution of permeability values, deviations are larger in real-scale values. All models are therefore generally considered to be robust (Table 1) and are interpreted to not be overfitting the training data.

3.2. Porosity prediction

Using a Random Forest model, porosity was predicted based on 60 features. The predicted porosity values based on petrographic input data show an overall good agreement with petrophysically determined values (Fig. 2). The R^2 of the measured versus predicted porosity for the test is high at 0.92, with a *MAE* of 1.25%, and a *RMSE* of 1.56%. These metrics imply a robust performance of the porosity prediction based on petrographic data. The errors are acceptable considering the performance in predicting porosity from four different regions, two different stratigraphic groups, and a porosity range from 0 to 20%. Furthermore, it can be observed that predictions for samples from all wells and regions are consistently plotting around the 1:1 line. This implies that the model does not over- or underpredict porosity as a function of the stratigraphic section or region/well.

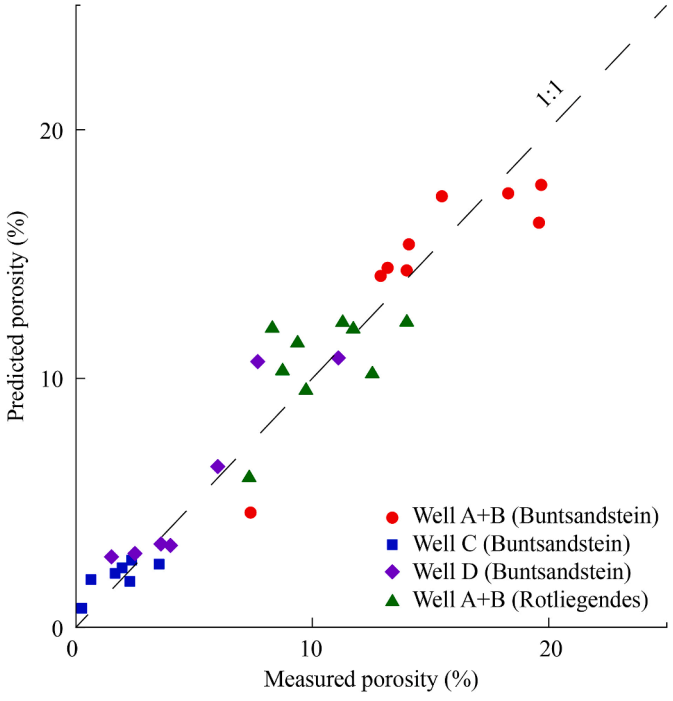

Fig. 2. Measured versus predicted porosity values for the 32 test samples from all studied wells. The dashed line represents the 1:1 correlation line.

3.3. Feature analysis for porosity prediction

SHAP dependence plots were generated based on the Random Forest model to analyze feature importance on porosity prediction (Fig. 3). The SHAP dependence plots illustrate the impact of individual features on porosity predictions, while also highlighting interactions of that feature with the most influential secondary feature. These interactions help reveal how two features jointly influence porosity, capturing complex dependencies beyond simple correlations.

Unsurprisingly, the intergranular porosity from point-counting analyses shows a positive non-linear relationship with porosity measurements on the associated plug samples (Fig. 3a). Positive SHAP values imply a positive influence on the target property. In this case, if the point-counted intergranular porosity is larger, the measured plug porosity is also larger. The non-linearity of this correlation can be related to the fact that optical porosity from point-counting analyses often underestimates microporosity (e.g., within clay minerals) (Hurst and Nadeau, 1995).

A location-specific control on porosity prediction was observed for authigenic titanium oxide contents (TiOx, rutile and anatase), which are mostly present in samples from well D (Buntsandstein). They show a positive correlation to porosity, although its contents are consistently below 4% (Fig. 3b). This location-specific relationship is only observed in samples from well D (Buntsandstein) and individual samples from well A + B (Buntsandstein), whereas the other sample series do not show elevated authigenic TiOx contents. Petrographic evaluation showed the TiOx appearing mostly in and around leached K-feldspar grains and surrounding expanded biotite flakes (Busch et al., 2025). Therefore, the authigenesis of TiOx is related to the initial presence of K-feldspar and biotite, which is partially dissolved and replaced by TiOx (Morad, 1986; Morad and Aldahan, 1987; Pe-Piper et al., 2011), thereby enhancing porosity. This correlation implies a local provenance-related effect, relating enhanced porosity to TiOx authigenesis and associated dissolution of detrital grains.

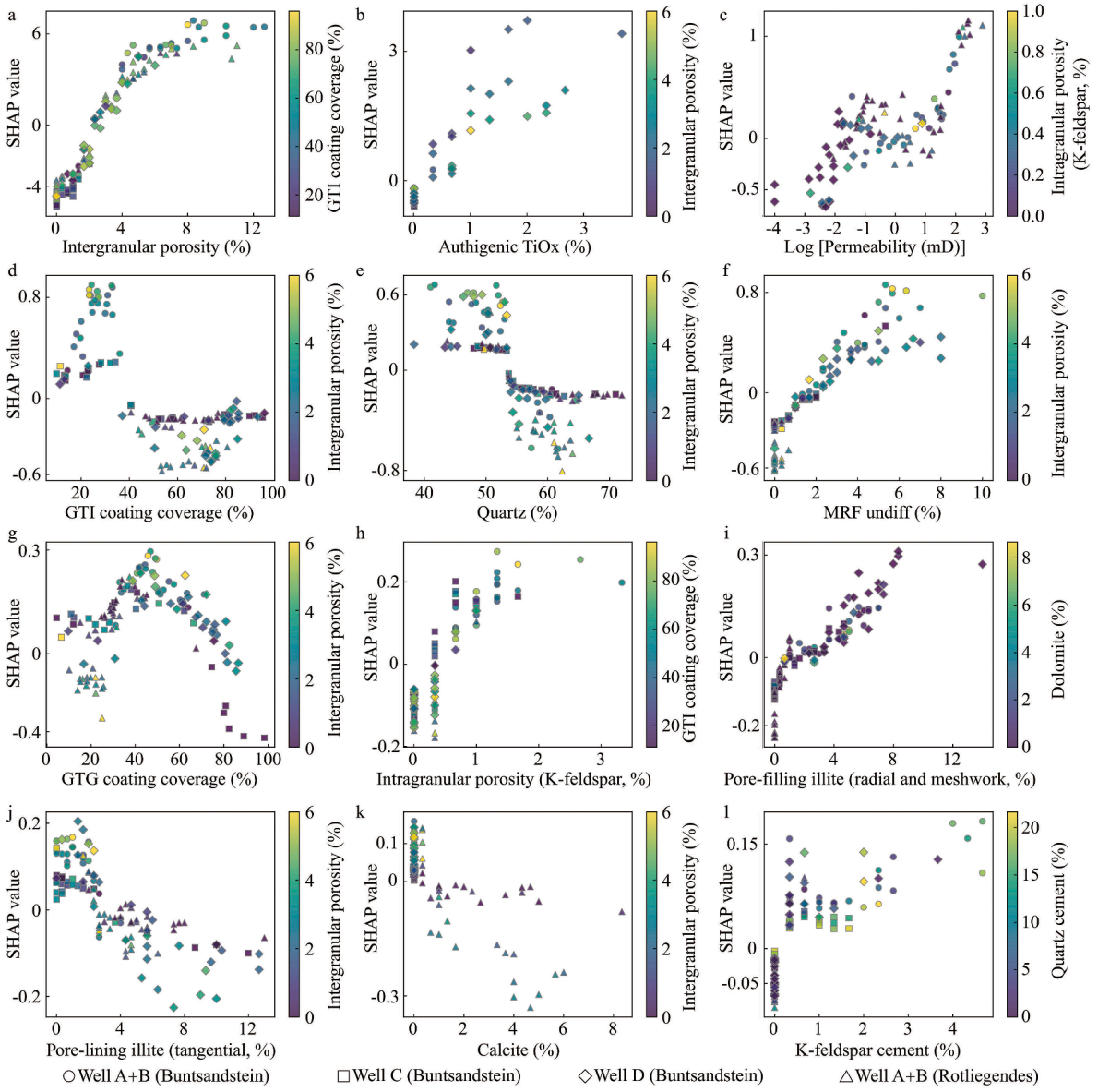

Fig. 3. SHAP dependence plots showing the influence of individual features (x-axis) on model predictions (y-axis), color-coded for the feature interactions: (a) intergranular porosity, (b) authigenic TiOx, (c) log(Permeability), (d) GTI coating coverage, (e) quartz grains, (f) undifferentiated MRF, (g) GTG coating coverage, (h) intragranular porosity in K-Feldspar grains, (i) pore-filling (radial and meshwork) illite, (j) pore-lining (tangential) illite, (k) calcite cement, and (l) K-feldspar cement.

Permeability and porosity do not exhibit a consistently strong positive correlation in the model (Fig. 3c). However, for samples with permeabilities above approximately 30 mD [lg(Permeability) = 1.4771], the model predicts a clearer positive relationship with porosity. This suggests that higher permeability correlates with higher porosity, implying generally larger pore throat diameters (i.e., better pore connectivity) at higher porosities. While permeability data shows a distinct influence on porosity predictions, it is not the strongest influence based on maximum SHAP values ~1. Porosity predictions are still more reliant on intergranular porosity from point-counting.

The GTI coating coverage influences porosity depending on its value range (Fig. 3d). If the GTI coating coverage is below 40%, the model predicts higher porosity values. However, when GTI coating coverage exceeds 40%, SHAP values become negative, meaning higher GTI coating coverage leads to lower porosity predictions. Generally, higher GTI coating coverages should reduce the surface area available for syntaxial quartz cement precipitation, thereby preserving porosity (Storvoll et al., 2002; Løvstad et al., 2022; Busch et al., 2024).

Peculiarly, most of the samples below 40% GTI are from well A + B (Buntsandstein), some from well C (Buntsandstein), and a few from well D (Buntsandstein), whereas all samples of well A + B (Rotliegendes) and most of well C + D (Buntsandstein) are above 40% GTI coating coverage. As most samples from well A + B (Buntsandstein) and some from well C + D (Buntsandstein) contain radial GTI coatings, but all samples from well A + B (Rotliegendes) and most of well C + D (Buntsandstein) contain tangential GTI grain coatings, the clay mineral texture affects porosity development. Notably, tangential GTI coatings are commonly present at GTG contacts and are known to enhance chemical compaction (i.e., pressure dissolution), reducing porosity (Heald, 1955; Thomson, 1959; Heald and Larese, 1974; Bjørkum, 1996; Kristiansen et al., 2011).

Interestingly, a similar trend is observed in the quartz grain content dependence plot, where detrital quartz grain content above 55% results in lower porosity predictions (Fig. 3e). This may indicate that high quartz content is associated with increased quartz overgrowth cementation, which reduces porosity. Alternatively, chemical compaction of detrital quartz grains only occurs if enough GTG contacts are formed in between quartz grains, a probability which is enhanced if more quartz grains are present in the studied samples. The latter option is favored, as a similar sample split is observed as in the previous analysis.

GTG coating coverages show a different pattern compared to GTI coating coverages (Fig. 3g). GTG values between 30% and 60% correspond to positive SHAP values, meaning that moderate grain contact coatings are beneficial for porosity preservation. However, when GTG coating coverages exceed 75%, the model predicts lower porosity. This is due to enhanced chemical compaction at illite coated grain contacts reducing the intergranular volume (Heald, 1955; Monsees et al., 2020). Chemical compaction enhanced by illite and muscovite is often described in literature (Heald, 1955) and has been related to differences in electrochemical surface potentials (Greene et al., 2009; Kristiansen et al., 2011).

Pore-filling illite (radial and meshwork textures) positively influences porosity, since this type of illite inhibits syntaxial quartz overgrowth cementation and preserves porosity (Fig. 3i). This aligns with diagenetic models where radial illite coatings limit quartz overgrowths, reducing overgrowth cement volumes (Busch et al., 2022). On the other hand, pore-lining (tangential) illite exceeding 3% decreases porosity (Fig. 3j), as it forms denser coatings and can also contribute to chemical compaction (pressure dissolution) of quartz grains if present at GTG contacts, while locally also inhibiting syntaxial quartz overgrowth cementation.

Furthermore, positive correlations of undifferentiated metamorphic rock fragment contents (MRF undiff, excluding phyllosilicate rich metamorphic rocks, Fig. 3f), K-feldspar cement contents (Fig. 3l), and intragranular porosity in K-feldspar grains were observed (Fig. 3h). Higher K-feldspar cement contents (almost exclusively affecting the Buntsandstein sample series) are positively affecting porosity predictions, likely related to their minor influence in stabilizing the grain framework against mechanical compaction. Unsurprisingly, higher intragranular porosity within partially dissolved K-feldspar grains also positively affects porosity predictions. The positive effect of elevated undifferentiated MRF content on porosity predictions was, at first glance, quite surprising, as normally, metamorphic rock fragments contain elevated sheet silicate contents. They are thus more liable to deform in a ductile manner, reducing the IGV and potentially porosity during compaction. However, as this category mostly encompasses meta-siltstones and meta-sandstones, their elevated quartz and feldspar content do not enhance the effect of compaction and thus preserve porosity.

A negative correlation was found for calcite cement, which was only observed in samples from well A + B (Rotliegendes) (Fig. 3k). Higher calcite contents therefore have a negative effect on porosity, by occluding pore spaces in the studied samples. Where calcite is absent, a slight positive influence on porosity predictions is recorded, as smaller pore-filling cement contents often relate to higher porosity, unless the cements are essential in stabilizing the grain framework against compaction.

Similarly, if contents of a certain property are zero, the SHAP values imply the opposite of the general correlation for that property. For example, if authigenic TiOx contents positively affect porosity predictions, their absence will slightly negatively influence porosity predictions (Fig. 3b). The same can be observed for undifferentiated MRF, intragranular porosity in K-feldspar grains, pore-filling illite, and K-feldspar cement (Figs. 3f, h, i, l).

Regarding feature interactions, higher contents of the secondary feature mostly result in more extreme influence on the predicted properties (Figs. 3d–g, j, k). For example, samples with a high GTI coating coverages (around 60%) and no intergranular porosity have a less negative influence on predicted porosity than a sample with the same GTI coating coverages and higher intergranular porosities (e.g., 2%–4%) (Fig. 3d). Some feature interactions show the opposite, with less extreme influences on model predictions if the secondary feature content is higher (Figs. 3b, l). For the remaining features, interactions with other features are less pronounced or absent (Figs. 3a, c, h, i).

3.4. Permeability prediction

The dataset for permeability prediction consists of 123 data points, with well C (Buntsandstein) excluded due to the absence of permeability measurements.

Permeability values in the studied sample suites exhibit a right-skewed, approximately log-normal distribution (Figs. 4a, b), spanning nearly seven orders of magnitude. Surprisingly, the best performing support vector machine model following hyperparameter optimization uses a linear kernel.

Real-scale permeability predictions using the optimized SVR model yield a R^2 of 0.81, a *MAE* of 29.4 mD and a *RMSE* of 68.4 mD, indicating that the model captures the overall trend but with substantial scatter at high permeability values. In real scale, cross validation results further show mean R^2 values of 0.50 $\pm$ 0.38, reflecting the impact of the extreme permeability range on fold-to-fold variability. These relatively high error values reflect the extreme range of permeability in the dataset, where large discrepancies in high-permeability predictions can disproportionately affect overall error metrics. To improve interpretability, a range-based *NRMSE* of 8.76% was also calculated. Log-transforming both measured and predicted permeability results in more stable and meaningful error estimates, with an R^2 of 0.83, an *MAE* of 0.21 and *RMSE* of 0.24 in log-space (Fig. 5). Given the approximately log-normal nature of the data, these values indicate that predictions generally fall within 0.21–0.24 orders of magnitude of the measured permeability values. Considering the data's broad dynamic range, these errors are acceptable, supporting the conclusion that permeability can be reasonably predicted within one order of magnitude, an appropriate scale for practical reservoir characterization. Similar to the porosity prediction, it can be observed that predictions for samples from all wells and regions are consistently plotting around the 1:1 line for the log-transformed permeability values (Fig. 5). This again implies that the model does not over- or under-predict permeability as a function of the stratigraphic section or region/well.

3.5. Feature analysis for permeability prediction

Due to the underlying linear kernel of the applied SVR, the SHAP dependency plots for the permeability prediction only show linear correlations (Fig. 6).

Petrophysically derived porosity and point-counted intergranular porosity show a strong correlation and a positive influence on the predicted permeability (Figs. 6a, d), as expected, as generally a positive correlation between porosity and permeability is often observed for sandstones. Similarly, intergranular porosity shows a positive influence on permeability prediction, and a correlation with pore-filling illite contents, whereas K-feldspar cement content shows a negative influence on permeability (Fig. 6e).

While the permeability prediction of the presented study relies strongly on petrophysical porosity data (SHAP values ranging from -0.6 to ~0.6) obtained from the studied samples, we argue that this

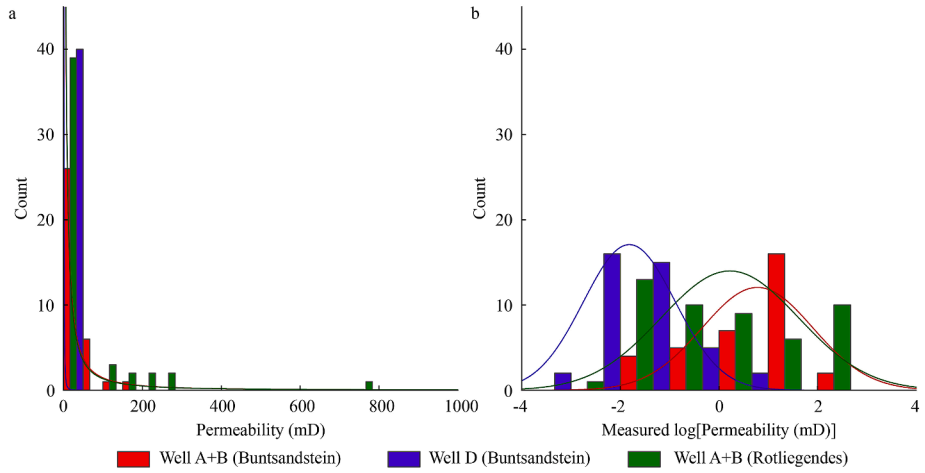

Fig. 4. Histograms showing the original permeability distribution (a) and the log-transformed permeability distribution (b).

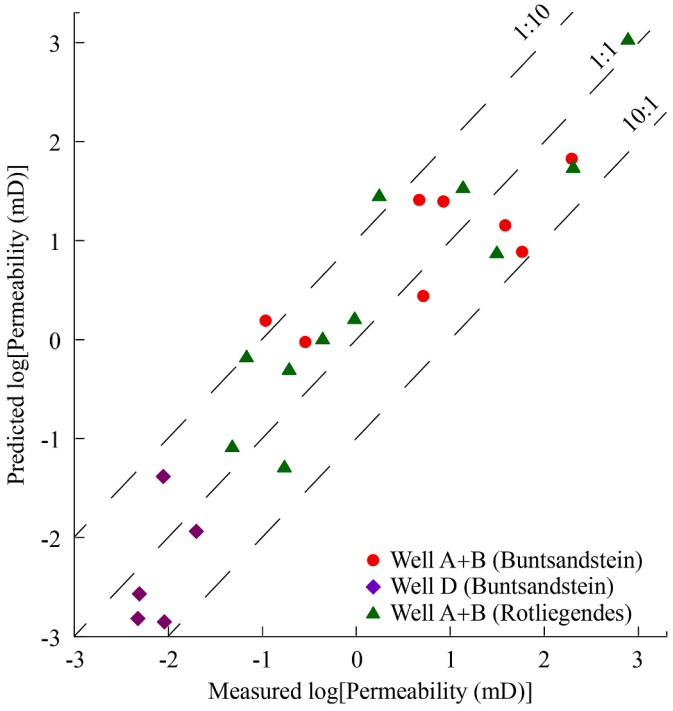

Fig. 5. Measured versus predicted log-transformed permeability values for the 24 test samples from the studied wells, except well C. The dashed lines represent the 1:10, 1:1, and 10:1 correlation lines.

data can also be derived based on measurements on other types of samples, from which petrographic samples can be prepared (e.g., cuttings or small rock chips). Unlike direct permeability measurements, porosity can be obtained from cuttings e.g., using standard laboratory porosity measurements relying on Archimedes' principle, mercury injection, μCT, or NMR measurements (Miritchnik et al., 2004; Hübner, 2014; Kesserwan et al., 2017; Chang et al., 2020). Therefore, the inclusion of reasonably obtainable porosity data from cuttings in addition to petrographic analyses will likely enhance the performance of machine learning predictions.

GTG coating coverages show a negative impact on permeability predictions, implying that their influence on chemical compaction also negatively affects permeability (Fig. 6b). The interaction with intergranular porosity suggests that regardless of intergranular porosity, an increase in GTG coating coverage leads to a permeability reduction. This suggests that chemical compaction plays a critical role in controlling fluid flow. Similarly, pore-lining illite (tangential), present as a grain coating mineral phase occupying GTG interfaces, negatively affects permeability predictions (Fig. 6g) due to its enhancement of chemical compaction of quartz grains.

Higher pore-filling illite contents lead to a reduced permeability prediction (Fig. 6c). This is in agreement with previous laboratory studies where radial and meshwork illite cementation reduces the effective pore throat radii, negatively affecting fluid flow (Neasham, 1977). However, samples containing higher pore-filling (radial and meshwork) illite contents show a positive interaction with intragranular porosity, relating to the inhibiting effects of radial illite on syntaxial quartz overgrowth cementation.

GTI coating coverage exhibits a generally positive relationship with permeability predictions, as indicated by the SHAP analysis (Fig. 6h). At lower GTI values (below ~50%), the SHAP values are slightly negative, suggesting a minor negative influence on predicted permeability. As GTI coverage increases, SHAP values rise steadily and become positive, reflecting a growing positive contribution to the model's predictions. (Fig. 6h). This is due to the inhibiting effects of grain coatings on syntaxial quartz overgrowth cementation (Bloch et al., 2002).

Unsurprisingly, grain size shows a positive influence on predicted permeability (Fig. 6f), as larger grain sizes preserve larger pore throat diameters, preserving permeability. Similarly, undifferentiated ductile rock fragments negatively affect permeability predictions (Fig. 6i), as they enhance the effects of mechanical compaction, reducing the IGV (Paxton et al., 2002).

Regarding feature interactions, higher contents of the secondary feature result in either larger or smaller influence on the predicted properties (Figs. 6a, c, d, f, i), while for the remaining features, interactions with other features are less pronounced or absent (Figs. 6b, e, g, h).

3.6. Comparison to linear regression baseline

The most important feature for both porosity and permeability predictions is the optical porosity or petrophysically derived porosity. Therefore, we also compare the performance of the best-performing machine learning models with linear regression calculations based on the correlation of porosity with permeability

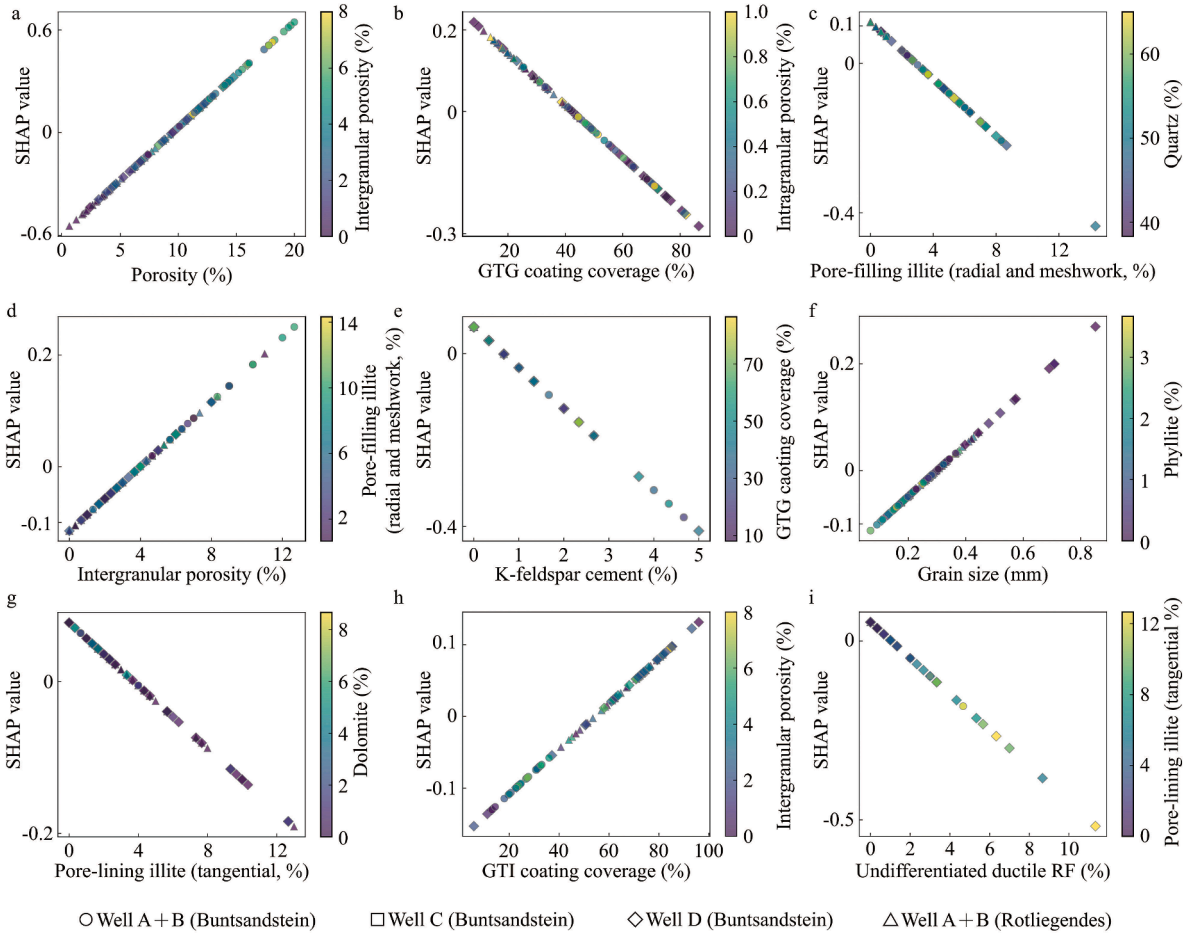

Fig. 6. SHAP dependence plots showing the influence of individual features (x-axis) on model predictions (y-axis), color-coded for the feature interactions: (a) porosity, (b) GTG coating coverage, (c) pore-filling (radial and meshwork) illite, (d) intragranular porosity, (e) K-feldspar cement, (f) grain size, (g) pore-lining (tangential) illite, (h) GTI coating coverage, and (i) undifferentiated ductile rock fragments (RF).

(log-transformed), optical porosity with permeability (log-transformed), and optical porosity with porosity.

The presented SVR model to predict permeability outperforms both the permeability calculated based on porosity or optical porosity (Table 2). Similarly, the porosity calculated based on optical porosity is outperformed by the presented RF model (Table 2). Therefore, the multivariate database used for the machine learning model setup consistently outperform linear regression calculations based on a single feature.

Table 2
Comparison of machine learning model metrics with linear regression calculations based on the porosity or optical porosity.

Machine learning				
Predictor	Feature(s)	R^2	*MAE*	*RMSE*
Porosity (RF)	See section 2.1.	0.92	1.25	1.56
Permeability (SVR)	See section 2.1.	0.83	0.21	0.24
Linear regression				
Porosity	Optical porosity	0.74	1.92	2.82
Permeability	Porosity	0.31	1.06	1.33
Permeability	Optical porosity	0.45	0.79	1.19

Notes: RF denotes Random Forest, SVR denotes Support Vector Regression.

4. Discussion

The presented approach successfully showcases the ability of machine learning approaches to derive meaningful porosity and permeability predictions regardless of present-day depths and diagenetic effects. Effects of diagenesis and compaction are recorded by the specific petrographic samples, as their authigenic mineral content, IGV, and the types and amounts of optically determined porosity differs.

The porosity model shows robust performance, reflected by high test metrics ($R^2 = 0.92$) and low cross-validation (CV) variability (mean CV $R^2 = 0.84$–0.86 with standard deviations of only 0.04–0.06). In contrast, real-scale permeability metrics show larger fold-to-fold variability (mean CV $R^2 = 0.32$–0.50 with high standard deviations), which is expected for permeability datasets that span several orders of magnitude. Importantly, when permeability is evaluated in log-space, consistent with its approximately log-normal distribution, the model performance becomes stable (CV $R^2 = 0.77 \pm 0.05$; test $R^2 = 0.83$) and error magnitudes remain low (*MAE* = 0.24 orders of magnitude). This demonstrates that the model captures permeability trends reliably when assessed in the appropriate statistical domain, even though real-scale predictions appear more variable due to data skewness rather than model instability.

Uncertainty in model predictions was evaluated using out-of-fold (OOF) cross-validation (Kohavi, 1995), which provides well-

Table 3
Per-well prediction uncertainties for permeability and porosity based on out-of-fold (OOF) cross-validation. Permeability errors (*MAE* and *RMSE*) are reported in millidarcies (mD), and porosity errors in percentage points. Ranges (minimum–maximum) represent the intrinsic variability of each well.

Wells	*N*	Porosity range (%)	*MAE* (%)	*RMSE* (%)	Permeability range (mD)	*MAE* (mD)	*RMSE* (mD)
Well A (Buntsandstein)	31	7.40–20.00	1.84	2.31	0.022–196.25	24.06	40.33
Well B (Buntsandstein)	3	13.90–14.30	0.98	1.14	21.50–38.50	16.46	18.89
Well C (Buntsandstein)	34	0.02–4.02	0.80	0.97	N/A	N/A	N/A
Well D (Buntsandstein)	40	2.39–15.95	1.98	2.38	0.0001–7.76	0.34	1.22
Well A (Rotliegend)	14	2.60–9.30	1.36	1.81	0.17–105.52	10.64	22.62
Well B (Rotliegend)	35	0.60–14.50	1.13	1.48	0.009–780.82	46.18	93.28

Note: N/A denotes non-available.

independent error estimates for all samples. Permeability uncertainty varies strongly between wells and closely follows the natural variability of the data (Table 3). For example, well B (Rotliegendes) exhibits the broadest permeability range (0.009–780 mD; $\sigma = 152$ mD), and correspondingly shows higher prediction errors (*MAE* = 46.2 mD; *RMSE* = 93.3 mD). In contrast, well D (Buntsandstein) displays a narrower permeability distribution (0.0001–7.76 mD; $\sigma = 1.23$ mD), and the model achieves very low uncertainty (*MAE* = 0.34 mD; *RMSE* = 1.22 mD). Wells with moderate permeability ranges, such as well A (Buntsandstein) (0.02–196 mD) and well A (Rotliegendes) (0.17–105 mD), exhibit proportionally moderate errors. Porosity predictions show uniformly low uncertainty across all wells, with porosity ranges typically between 0.02% and 20% and per-well *MAE* values of only 0.80%–1.98%. These results demonstrate that prediction uncertainty is primarily governed by the intrinsic geological variability of each well rather than by model instability. Importantly, no well shows systematic degradation in predictive performance, indicating that the models generalize robustly across all six wells and both reservoir systems.

Furthermore, model robustness is supported by their ability to assess porosity and permeability for different sample suites from four regions in northern and southern Germany, from the Permian and Triassic, both forming major reservoir systems in central Europe (Table 3). Although samples in both regions were deposited in arid fluvio-eolian systems, samples from the Rotliegendes mostly represent eolian depositional environments (Gast et al., 2010; Monsees et al., 2020), whereas samples from the Buntsandstein mostly represent fluvial deposits (Bourquin et al., 2009) with occasional eolian intercalations (Quandt et al., 2022). Nonetheless, porosity and permeability predictions based on this model do not appear to be affected by these differences in depositional environments, as only the detrital and authigenic components from point-counting analyses are considered. Even the local abundance of certain authigenic minerals [TiOx in well D (Buntsandstein), calcite in well A + B (Rotliegendes)] does not pose a problem, as the absence of these minerals in other samples does not strongly affect porosity or permeability predictions based on SHAP plot evaluation. The differences in detrital composition (e.g., higher volcanic rock fragment and chert contents in the Rotliegendes samples) and burial histories did not affect model performance, as the mix of detrital and authigenic phases and their respective volumes both affect compaction and porosity. Therefore, porosity and permeability predictions for specific petrographic sample compositions are possible.

4.1. Limitations and challenges

Currently perceived challenges include operator bias in petrographic analyses, as available datasets span multiple decades and providers. Operator bias was excluded in this study, as all data was prepared by petrographers from the same institution, resulting in a limited number of samples of 157. Future applications should likely account for bias from different petrographers and evolving point-counting categories. Enhancing machine learning-based classifications of mineral and rock types may help overcome this bias if mineral textures are adequately differentiated, as e.g., textures of clay mineral cements matter for permeability (Neasham, 1977). This includes the accurate differentiation of rock fragments (as also highlighted by Tang et al., 2020), and accurate classification of detrital grains and syntaxial overgrowth cements showing the same optical properties. Furthermore, due to the limited number of samples from just two stratigraphic units, not all petrographic characteristics, that can generally be observed in sandstones, are captured. The generalizability of the presented models has to be improved in future applications by including more stratigraphic units and more diverse detrital and authigenic compositions.

Scientific advancements may change point-counting categories, affecting database detail. For example, grain coating clay minerals impact quartz cement inhibition (Heald and Larese, 1974) and chemical compaction (Heald, 1955) but are rarely quantified (i.e., grain coating coverages at different interfaces, Monsees et al., 2020). This work is limited to certain petrographic datasets with controlled point-counting categories and quantified diagenetic reservoir quality controls. Generally, machine learning approaches can adapt to evolving scientific understanding and novel quantifications based on petrographic samples. This assessment is supported by the location-specific presence of certain minerals, which, if present, affect the predictions, but their absence does not necessarily strongly affect model predictions (Fig. 3).

A limitation of this study is the use of shuffled data, which prevents models from capturing depth-dependent geological relationships, which are documented in a variety of geological settings such as fining- or coarsening upwards sequences. The small dataset, consisting of only six wells (four regions) and permeability data from just five wells (three regions), restricted the possibility of well-wise splitting, limiting the model's ability to recognize formation-specific trends. However, SHAP plots indicate a good performance in mixed sample series. Additionally, core sampling gaps with irregular intervals resulted in variable sampling frequency, making it difficult to establish a consistent spatial representation of geological features with depth. This irregularity complicates data augmentation and could pose challenges for ANN-based approaches, as these methods rely on structured spatial patterns.

If more data were available, well-wise splits could allow models to learn location-based trends and depositional variations, but this does not necessarily guarantee better generalization. Instead, a blind-well testing approach would provide a more realistic validation of model performance by assessing its ability to predict porosity and permeability of unseen wells. Future studies should consider combining well-wise splits with blind-well validation to balance both pattern-learning and generalizability in machine learning-driven reservoir quality assessment and prediction.

5. Future application potential

If a machine learning model is suitably calibrated to available core material that captures the range of effects of diagenetic alteration and compaction, it could be used to estimate likely porosity and permeability ranges based on petrographic data (i.e., point-counting analyses and granulometry) and data derived from these analyses. This may be especially valuable in scenarios, where due to slim economic margins (geothermal, subsurface gas storage, or brown-field extension) no core material is taken or where core was lost, and routine core analyses cannot be performed. An application could be to use cuttings to prepare petrographic samples and apply appropriate point-counting analyses (while skipping the area outside of individual cuttings in the thin sections, Ölmez et al., 2025) and feed the results into a model trained on that lithology, given the cuttings are of a suitable quality and contain a mostly undisturbed rock fabric. Since cuttings are available for extended well sections, their results may also provide a more detailed assessment on diagenetic reservoir quality controls (Radwan, 2022), while potentially being used to gain insights into porosity and permeability. For example, direct porosity measurements can be readily obtained from cuttings or small rock chips, and the performance of permeability predictions can be improved. Thereby, meaningful porosity and permeability ranges may be estimated for well sections with no core material available, given that suitable core material of the target formation is available for training.

Geological datasets, including petrographic and core analysis data, are available for some wells in countries like the Netherlands and Norway (e.g., nlog or DISKOS), providing a basis for further tests. This can help create robust machine learning models to assess petrophysical properties of uncored wells, which is important for optimal use of subsurface samples and can be applied to cuttings as well, for which databases already exist (Norwegian Offshore Directorate, 2022).

An added benefit of the approach is the possibility to assess detrital and authigenic reservoir quality controls based on SHAP values (Figs. 3 and 6), to enhance the understanding of reservoir systems. The provided SHAP values outline the petrographic properties that have a positive or negative influence on the target variable (here: porosity and permeability) and may be used as a tool to better understand reservoir quality controls that are related to diagenesis and compaction. For this dataset, the previously published reservoir quality controlling processes (enhanced chemical compaction along illite coated grain contacts, illite coatings inhibiting quartz cementation and preserving some porosity, radial illite coatings reducing permeability) were also uncovered using machine learning approaches by assessing SHAP plots. While these correlations are revealed by machine learning approaches, they still require geological explanations to derive the underlying processes. This may be an addition to the standard workflow of manually assessing correlations between properties and performing rock typing analyses, which can be liable to missing relevant controlling factors. This enhanced understanding will ultimately allow better pre-drill reservoir quality predictions, as relevant controlling factors on reservoir quality development are captured. These controlling factors can then be applied to predictive reservoir quality modeling including the effects of pressure and temperature on the rate of compaction, cementation, and dissolution (Lander and Walderhaug, 1999; Walderhaug et al., 2000; Makowitz et al., 2006; Busch et al., 2018; Tamburelli et al., 2022).

In addition to previous studies, this approach further highlights the flexibility of machine learning approaches when applied to geological data. It is therefore likely that further datasets, for example geochemical datasets (e.g., from XRF/pXRF/μXRF, EDX) or mineralogical datasets (e.g., XRD, mineral mapping), can be easily integrated and utilized to maximize the knowledge gain from subsurface samples and predict reservoir-relevant data both in regression and classification tasks.

Machine learning approaches may thus help assess porosity and permeability ranges using standard petrographic samples of cuttings when subsurface core material is available for training. Furthermore, if suitable calibration data is available from outcrops, this approach may also be used to gain porosity and permeability ranges based on petrographic samples prepared from outcrop material. However, without sufficient core material, routine core analysis, and accurate petrographic reservoir quality analyses, predictions may be inaccurate due to incorrect assessment of controlling factors or local variations in diagenetic alteration. Literature documents local variability within formations, such as differences in authigenic clay minerals (e.g., illite, chlorite or kaolinite coatings in the Rotliegendes) (Waldmann and Gaupp, 2016; Molenaar and Felder, 2018; Busch et al., 2020) or lithofacies changes (e.g., marginal and basinal facies in the Buntsandstein) (Olivarius et al., 2015; Quandt et al., 2022), affecting porosity-permeability relationships. Accurate predictions of models like these require calibration to authigenic and detrital phases and sandstone textures, showing an influence on petrophysical properties, emphasizing the need for geological understanding and calibration using core material.

6. Conclusions

The presented approach well captures the relevant petrographic factors influencing porosity and permeability in the studied samples. Their interactions, using geological reasoning, can be used to assess reservoir quality controls derived from SHAP plots, to enhance the understanding of reservoir systems.

The robustness of the approach in predicting porosity and permeability in Permian Rotliegendes and Triassic Buntsandstein samples could be demonstrated. Even only locally occurring mineral species (calcite in the Rotliegendes and TiOx in the Buntsandstein) can be accurately captured and included in predictions. This implies the necessity of a trained model to capture the inherent mineralogical variety in diagenetically overprinted sedimentary rocks to make accurate predictions.

Generally, the porosity prediction is very good with a $R^2 = 0.92$, a *MAE* of 1.25%, and a *RMSE* of 1.56%. The permeability prediction of real-scale permeability is good with a $R^2 = 0.85$, a *MAE* of 29.4 mD, a *RMSE* of 68.3 mD, and a range-based normalized *RMSE* of 8.76%. Log-transformation of the measured and predicted permeability values delivers a more representative R^2 of 0.83, *MAE* of 0.21 and a *RMSE* of 0.24. Given the log-normal distribution of permeability datasets, obtaining permeability predictions with a range-based normalized *RMSE* (real-scale) of 8.76% or a *RMSE* (log-transformed) of ±0.24 orders of magnitude is acceptable, considering the alternative of no indication of either porosity or permeability.

Following the promising initial proof of concept, this approach can be extended to larger datasets, incorporating blind-well validation and utilization of legacy datasets prepared by various petrographers. A potential application of this approach is to gain porosity and permeability ranges from e.g., cuttings in uncored wells, providing samples that can still be used for petrographic analyses, provided that nearby calibration data is available.

CRediT authorship contribution statement

Sahar Sadrikhanloo: Writing – review & editing, Methodology, Investigation, Formal analysis. **Benjamin Busch:** Writing – original

draft, Methodology, Investigation, Conceptualization. **Christoph Hilgers:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Acknowledgements

BB and CH thankfully acknowledge previous provision of sample material by operators. The authors thankfully acknowledge constructive review comments by Assoc. Editor Li Dong and two anonymous reviewers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.engeos.2026.100537.

References

Abbas, M.A., Al-Mudhafar, W.J., Wood, D.A., 2023. Improving permeability prediction in carbonate reservoirs through gradient boosting hyperparameter tuning. Earth Sci. Inform. 16 (4), 3417–3432. https://doi.org/10.1007/s12145-023-01099-0.

Ajdukiewicz, J.M., Lander, R.H., 2010. Sandstone reservoir quality prediction: the state of the art. AAPG (Am. Assoc. Pet. Geol.) Bull. 94 (8), 1083–1091. https://doi.org/10.1306/intro060110.

Bergen, K.J., Johnson, P.A., de Hoop, M.V., Beroza, G.C., 2019. Machine learning for data-driven discovery in solid Earth geoscience. Science 363, 6433. https://doi.org/10.1126/science.aau0323.

Bjørkum, P.A., 1996. How important is pressure in causing dissolution of quartz in sandstones? J. Sediment. Res. 66 (1), 147–154.

Bloch, S., Lander, R.H., Bonnell, L.M., 2002. Anomalously high porosity and permeability in deeply buried sandstone reservoirs: origin and predictability. AAPG (Am. Assoc. Pet. Geol.) Bull. 86, 301–328. https://doi.org/10.1306/61EEDABC-173E-11D7-8645000102C1865D.

Bourquin, S., Guillocheau, F., Péron, S., 2009. Braided rivers within an arid alluvial plain (example from the Lower Triassic, western German Basin): recognition criteria and expression of stratigraphic cycles. Sedimentology 56 (7), 2235–2264. https://doi.org/10.1111/j.1365-3091.2009.01078.x.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brochu, E., Cora, V.M., de Freitas, N., 2010. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning.

Busch, B., Adelmann, D., Herrmann, R., Hilgers, C., 2022. Controls on compactional behavior and reservoir quality in a Triassic Buntsandstein reservoir, Upper Rhine Graben, SW Germany. Mar. Petrol. Geol. 136, 105437. https://doi.org/10.1016/j.marpetgeo.2021.105437.

Busch, B., Böcker, J., Hilgers, C., 2024. Improved reservoir quality assessment by evaluating illite grain coatings, quartz cementation, and compaction – case study from the Buntsandstein, Upper Rhine Graben. Germany Geoenergy Sci. Eng. 241, 213141. https://doi.org/10.1016/j.geoen.2024.213141.

Busch, B., Böke, L.L., Schmidt, C., Warnecke, M., Hilgers, C., 2025. Reservoir quality controls in the Lower Buntsandstein and their relation to detrital grain size, URG, SW-Germany. Geothermics 132, 103453. https://doi.org/10.1016/j.geothermics.2025.103453.

Busch, B., Hilgers, C., Adelmann, D., 2020. Reservoir quality controls on Rotliegend fluvio-aeolian wells in Germany and the Netherlands, Southern Permian Basin – impact of grain coatings and cements. Mar. Petrol. Geol. 112, 104075. https://doi.org/10.1016/j.marpetgeo.2019.104075.

Busch, B., Hilgers, C., Lander, R.H., Bonnell, L.M., Adelmann, D., 2018. Reservoir quality and burial model evaluation by kinetic quartz and illite cementation modeling: case study Rotliegendes, N-Germany. AAPG (Am. Assoc. Pet. Geol.) Bull. 101 (2), 293–307. https://doi.org/10.1306/0503171605217075.

Chang, Y., Yao, Y., Liu, Y., Zheng, S., 2020. Can cuttings replace cores for porosity and pore size distribution analyses of coal? Int. J. Coal Geol. 227. https://doi.org/10.1016/j.coal.2020.103534.

Cracknell, M.J., Reading, A.M., 2013. The upside of uncertainty: identification of lithology contact zones from airborne geophysics and satellite data using Random Forests and Support Vector Machines. Geophysics 78 (3), WB113–WB126. https://doi.org/10.1190/geo2012-0411.1.

Dramsch, J.S., 2020. 70 years of machine learning in geoscience in review. Machine Learn. Geosci. 1–55. https://doi.org/10.1016/bs.agph.2020.08.002.

Duarte, D., Pires de Lima, R., Tellez, J., Pranter, M.J., 2023. Spatial variability of petrofacies using supervised machine learning and geostatistical modeling: Sycamore formation, Sho-Vel-Tum field, Oklahoma, USA. Interpretation 11 (2), T289–T302. https://doi.org/10.1190/int-2022-0064.1.

Dubois, M.K., Bohling, G.C., Chakrabarti, S., 2007. Comparison of four approaches to a rock facies classification problem. Comput. Geosci. 33 (5), 599–617. https://doi.org/10.1016/j.cageo.2006.08.011.

Esmaeili, B., Hosseinzadeh, S., Kadkhodaie, A., Wood, D.A., Akbarzadeh, S., 2024. Simulating reservoir capillary pressure curves using image processing and classification machine learning algorithms applied to petrographic thin sections. J. Afr. Earth Sci. 209. https://doi.org/10.1016/j.jafrearsci.2023.105098.

Folk, R.L., 1980. Petrology of Sedimentary Rocks. Hemphill Publishing Company, Austin, Texas, U.S.A., p. 182

Gast, R.E., Dusar, M., Breitkreuz, C., Gaupp, R., Schneider, J.W., Stemmerik, L., Geluk, M.C., Geißler, M., Kiersnowski, H., Glennie, K.W., Kabel, S., Jones, N.S., 2010. Rotliegend. In: Doornenbal, J.C., Stevenson, A.G. (Eds.), Petroleum Geological Atlas of the Southern Permian Basin Area. EAGE Publications BV, Houten, pp. 101–123.

Greene, G.W., Kristiansen, K., Meyer, E.E., Boles, J.R., Israelachvili, J.N., 2009. Role of electrochemical reactions in pressure solution. Geochem. Cosmochim. Acta 73, 2862–2874. https://doi.org/10.1016/j.gca.2009.02.012.

Guido, R., Ferrisi, S., Lofaro, D., Conforti, D., 2024. An overview on the advancements of support vector machine models in healthcare applications: a review. Information 15, 4. https://doi.org/10.3390/info15040235.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46, 389–422. https://doi.org/10.1023/A:1012487302797.

Heald, M.T., 1955. Stylolites in sandstone. J. Geol. 63, 101–114. https://doi.org/10.1086/626237.

Heald, M.T., Larese, R.E., 1974. Influence of coatings on quartz cementation. J. Sediment. Petrol. 44 (4), 1269–1274. https://doi.org/10.1306/212F6C94-2B24-11D7-8648000102C1865D.

Hübner, W., 2014. Studying the pore space of cuttings by NMR and μCT. J. Appl. Geophys. 104, 97–105. https://doi.org/10.1016/j.jappgeo.2014.02.016.

Hurst, A., Nadeau, P.H., 1995. Clay microporosity in reservoir sandstones: an application of quantitative electron microscopy in petrophysical evaluation. AAPG (Am. Assoc. Pet. Geol.) Bull. 79 (4), 563–573.

Johnson, J.I., Sharman, G.R., Szymanski, E., Huang, X., 2022. Machine learning applied to a modern-pleistocene petrographic data set: the global prediction of sand modal composition (GloPrSM) model. J. Geophys. Res.: Earth Surf. 127 (7). https://doi.org/10.1029/2022jf006595.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H.A., Kumar, V., 2019. Machine learning for the geosciences: challenges and opportunities. IEEE Trans. Knowl. Data Eng. 31 (8), 1544–1554. https://doi.org/10.1109/tkde.2018.2861006.

Kesserwan, H., Alshanqaiti, E., Jin, G., Nuic, I., 2017. Porosity measurements on drill cuttings - comprehensive inputs to formation evaluation near real-time while drilling. In: Abu Dhabi International Petroleum Exhibition & Conference. SPE, Abu Dhabi, UAE.

Kohavi, R., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: International Joint Conference on Artificial Intelligence, pp. 1137–1143. Montreal, Quebec, Canada.

Kristiansen, K., Valtiner, M., Greene, G.W., Boles, J.R., Israelachvili, J.N., 2011. Pressure solution - the importance of the electrochemical surface potential. Geochem. Cosmochim. Acta 75, 6882–6892. https://doi.org/10.1016/j.gca.2011.09.019.

Lander, R.H., Walderhaug, O., 1999. Predicting porosity through simulating sandstone compaction and quartz cementation. AAPG (Am. Assoc. Pet. Geol.) Bull. 83 (3), 433–449. https://doi.org/10.1306/00AA9BC4-1730-11D7-8645000102C1865D.

Løvstad, K., Hansen, H.N., Jahren, J., 2022. The porosity preserving effect of basin wide illitic coating in deeply buried sandstone intervals of the lower Jurassic Stø Formation, Barents Sea. Mar. Petrol. Geol. 137. https://doi.org/10.1016/j.marpetgeo.2021.105498.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: von Luxburg, U., Guyon, I., Bengio, S., Wallach, H., Fergus, R. (Eds.), NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, Long Beach, CA, USA, pp. 4768–4777.

Makowitz, A., Lander, R.H., Milliken, K.L., 2006. Diagenetic modeling to assess the relative timing of quartz cementation and brittle grain processes during compaction. AAPG (Am. Assoc. Pet. Geol.) Bull. 90 (6), 873–885. https://doi.org/10.1306/12190505044.

Male, F., Duncan, I.J., 2020. Lessons for machine learning from the analysis of porosity-permeability transforms for carbonate reservoirs. J. Petrol. Sci. Eng. 187. https://doi.org/10.1016/j.petrol.2019.106825.

Miritchnik, K., Kryuchkov, S., Strack, K., 2004. A Novel Method to Determine Nmr Petrophysical Parameters from Drill Cuttings SPWLA 45th Annual Logging Symposium. SPWLA, Noordwijk, Netherlands.

Mishra, A., Sharma, A., Patidar, A.K., 2022. Evaluation and development of a predictive model for geophysical well log data analysis and reservoir characterization: machine learning applications to lithology prediction. Nat. Resour. Res. 31 (6), 3195–3222. https://doi.org/10.1007/s11053-022-10121-z.

Molenaar, N., Felder, M., 2018. Clay cutans and the origin of illite rim cement: an example from the siliciclastic Rotliegend Sandstone in the Dutch Southern Permian Basin. J. Sediment. Res. 88 (5), 641–658. https://doi.org/10.2110/jsr.2018.33.

Monsees, A.C., Busch, B., Schöner, N., Hilgers, C., 2020. Rock typing of diagenetically induced heterogeneities – a case study from a deeply-buried clastic Rotliegend reservoir of the Northern German Basin. Mar. Petrol. Geol. 113, 104163. https://doi.org/10.1016/j.marpetgeo.2019.104163.

Morad, S., 1986. SEM study of authigenic rutile, anatase and brookite in Proterozoic sandstones from Sweden. Sediment. Geol. 46, 77–89.

Morad, S., Aldahan, A.A., 1987. Diagenetic "replacement" of feldspars by titanium oxides in sandstones. Sediment. Geol. 51 (3–4), 147–153. https://doi.org/10.1016/0037-0738(87)90045-5.

Neasham, J.W., 1977. The morphology of dispersed clay in sandstone reservoirs and its effect on sandstone shaliness, pore space and fluid flow properties. In: SPE Annual Fall Technical Conference and Exhibition. Denver, CO.

Norwegian Offshore Directorate, 2022. NOROG digital cuttings Project. https://www.sodir.no/en/diskos/wells/historical-projects/norog-digital-cuttings-project/.

Olivarius, M., Weibel, R., Hjuler, M.L., Kristensen, L., Mathiesen, A., Nielsen, L.H., Kjøller, C., 2015. Diagenetic effects on porosity–permeability relationships in red beds of the Lower Triassic Bunter Sandstone Formation in the North German Basin. Sediment. Geol. 321, 139–153. https://doi.org/10.1016/j.sedgeo.2015.03.003.

Ölmez, J.A., Busch, B., Möbius, R., Dasgupta, K., Gauer, A., Tosoratti, F., Hilgers, C., 2025. Reservoir characterization and well production proxy analyses on drill cuttings: case study from the flysch play in the Vienna Basin (NE Austria). Energy Geosci. 6, 100461. https://doi.org/10.1016/j.engeos.2025.100461.

Ore, T., Gao, D., 2023. Prediction of reservoir brittleness from geophysical logs using machine learning algorithms. Comput. Geosci. 171. https://doi.org/10.1016/j.cageo.2022.105266.

Paxton, S.T., Szabo, J.O., Ajdukiewicz, J.M., Klimentidis, R.E., 2002. Construction of an intergranular compaction curve for evaluating and predicting compaction and porosity loss in rigid grain sandstone reservoirs. AAPG (Am. Assoc. Pet. Geol.) Bull. 86 (12), 2047–2067. https://doi.org/10.1306/61EEDDFA-173E-11D7-8645000102C1865D.

Pe-Piper, G., Karim, A., Piper, D.J.W., 2011. Authigenesis of titania minerals and the mobility of Ti: new evidence from pro-deltaic sandstones, Cretaceous Scotian Basin, Canada. J. Sediment. Res. 81 (10), 762–773. https://doi.org/10.2110/jsr.2011.63.

Pittman, E.D., Larese, R.E., Heald, M.T., 1992. Clay coats : occurrence and relevance to preservation of porosity in sandstones. In: Houseknecht, D.W., Pittmann, E.D. (Eds.), Origin, Diagenesis, and Petrophysics of Clay Minerals in Sandstones. SEPM Special Publication, Tulsa, Oklahoma, U.S.A., pp. 241–256. https://doi.org/10.2110/pec.92.47.0241

Quandt, D., Busch, B., Schmidt, C., Hilgers, C., 2022. Diagenesis and controls on reservoir quality of Lower Triassic red bed sandstones (Buntsandstein) from a marginal basin facies, Southwest Germany. Mar. Petrol. Geol. 142, 105744. https://doi.org/10.1016/j.marpetgeo.2022.105744.

Radwan, A.E., 2022. Provenance, depositional facies, and diagenesis controls on reservoir characteristics of the middle Miocene Tidal sandstones, Gulf of Suez Rift Basin: integration of petrographic analysis and gamma-ray log patterns. Environ. Earth Sci. 81 (15). https://doi.org/10.1007/s12665-022-10502-w.

Rahimi, M., Riahi, M.A., 2022. Reservoir facies classification based on random forest and geostatistics methods in an offshore oilfield. J. Appl. Geophys. 201. https://doi.org/10.1016/j.jappgeo.2022.104640.

Rider, M., Kennedy, M., 2011. The Geological Interpretation of Well Logs, 3 ed. Rider-French Consulting Ltd, Scotland.

Rubo, R.A., de Carvalho Carneiro, C., Michelon, M.F., Gioria, R.d.S., 2019. Digital petrography: mineralogy and porosity identification using machine learning algorithms in petrographic thin section images. J. Petrol. Sci. Eng. 183. https://doi.org/10.1016/j.petrol.2019.106382.

Saxena, N., Day-Stirrat, R.J., Hows, A., Hofmann, R., 2021. Application of deep learning for semantic segmentation of sandstone thin sections. Comput. Geosci. 152. https://doi.org/10.1016/j.cageo.2021.104778.

scikit-learn User Guide, 2025. 1.4. support vector machines. https://scikit-learn.org/stable/modules/svm.html.

Shahnas, M.H., Yuen, D.A., Pysklywec, R.N., 2018. Inverse problems in geodynamics using machine learning algorithms. J. Geophys. Res. Solid Earth 123 (1), 296–310. https://doi.org/10.1002/2017jb014846.

Storvoll, V., Bjørlykke, K., Karlsen, D., Saigal, G., 2002. Porosity preservation in reservoir sandstones due to grain-coating illite: a study of the Jurassic garn formation from the kristin and lavrans fields, offshore mid-norway. Mar. Petrol. Geol. 19, 767–781. https://doi.org/10.1016/S0264-8172(02)00035-1.

Sun, J., Zhang, R., Chen, M., Chen, B., Wang, X., Li, Q., Ren, L., 2021. Identification of porosity and permeability while drilling based on machine learning. Arabian J. Sci. Eng. 46 (7), 7031–7045. https://doi.org/10.1007/s13369-021-05432-x.

Tamburelli, S., Di Giulio, A., Amadori, C., Consonni, A., Ortenzi, A., 2022. New constraint on burial and thermal history of Devonian reservoir sandstones in the Illizi-Ghadames basin (North Africa) through diagenetic numerical modelling. Mar. Petrol. Geol. 145. https://doi.org/10.1016/j.marpetgeo.2022.105903.

Tang, D.G., Milliken, K.L., Spikes, K.T., 2020. Machine learning for point counting and segmentation of arenite in thin section. Mar. Petrol. Geol. 120. https://doi.org/10.1016/j.marpetgeo.2020.104518.

Thomson, A., 1959. Pressure solution and porosity. In: Ireland, H.A. (Ed.), Silica in Sediments. https://doi.org/10.2110/pec.59.01.0092.

Walderhaug, O., Lander, R.H., Bjørkum, P.A., Oelkers, H., Bjørlykke, K., Nadeau, P.H., 2000. Modelling quartz cementation and porosity in reservoir sandstones: examples from the Norwegian continental shelf. In: Worden, R.H., Morad, S. (Eds.), Quartz Cementation in Sandstones. Blackwell Science Ltd., Oxford, pp. 39–49. https://doi.org/10.1002/9781444304237.ch3.

Waldmann, S., Gaupp, R., 2016. Grain-rimming kaolinite in Permian Rotliegend reservoir rocks. Sediment. Geol. 335, 17–33. https://doi.org/10.1016/j.sedgeo.2016.01.016.

Wood, D.A., 2020. Predicting porosity, permeability and water saturation applying an optimized nearest-neighbour, machine-learning and data-mining network of well-log data. J. Petrol. Sci. Eng. 184. https://doi.org/10.1016/j.petrol.2019.106587.

Worden, R.H., Armitage, P.J., Butcher, A.R., Churchill, J.M., Csoma, A.E., Hollis, C., Lander, R.H., Omma, J.E., 2018. Petroleum reservoir quality prediction: overview and contrasting approaches from sandstone and carbonate communities. In: Armitage, P.J., Butcher, A.R., Churchill, J.M., Csoma, A.E., Hollis, C., Lander, R.H., Omma, J.E., Worden, R.H. (Eds.), Reservoir Quality of Clastic and Carbonate Rocks: Analysis, Modelling and Prediction. Geological Society, London, pp. 1–31. https://doi.org/10.1144/SP435.21.

Wrona, T., Pan, I., Gawthorpe, R.L., Fossen, H., 2018. Seismic facies analysis using machine learning. Geophysics 83 (5), O83–O95. https://doi.org/10.1190/geo2017-0595.1.

Yang, T., Tang, H., Dai, J., Wang, H., Wen, X., Wang, M., He, J., Wang, M., 2025. Quantitative classification and prediction of pore structure in low porosity and low permeability sandstone: a machine learning approach. Geoenergy Sci. Eng. 247. https://doi.org/10.1016/j.geoen.2025.213708.

Yu, J., Wellmann, F., Virgo, S., von Domarus, M., Jiang, M., Schmatz, J., Leibe, B., 2023. Superpixel segmentations for thin sections: evaluation of methods to enable the generation of machine learning training data sets. Comput. Geosci. 170. https://doi.org/10.1016/j.cageo.2022.105232.

Zanganeh Kamali, M., Davoodi, S., Ghorbani, H., Wood, D.A., Mohamadian, N., Lajmorak, S., Rukavishnikov, V.S., Taherizade, F., Band, S.S., 2022. Permeability prediction of heterogeneous carbonate gas condensate reservoirs applying group method of data handling. Mar. Petrol. Geol. 139. https://doi.org/10.1016/j.marpetgeo.2022.105597.