Machine Learning to Predict Effective Reaction Rates in 3D Porous Media From Pore Structural Features

Min Liu

University of Minnesota

Beomjin Kwon

Arizona State University

Peter K. Kang (✉ pkkang@umn.edu)

University of Minnesota

Research Article

Keywords: Effective reaction rate, Machine learning, Pore-scale modeling, Porous media, Reactive transport, Fluid-solid reaction

Posted Date: January 28th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1284059/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on March 31st, 2022. See the published version at <https://doi.org/10.1038/s41598-022-09495-0>.

Machine Learning to Predict Effective Reaction Rates in 3D Porous Media from Pore Structural Features

Min Liu^{1,2}, Beomjin Kwon³, Peter K. Kang^{1,2,*}

1. Department of Earth and Environmental Sciences, University of Minnesota, Minneapolis, USA

2. Saint Anthony Falls Laboratory, University of Minnesota, Minneapolis, USA

3. School for Engineering of Matter, Transport and Energy, Arizona State University, Phoenix, USA

Abstract

Large discrepancies between well-mixed reaction rates and effective reactions rates estimated under fluid flow conditions have been a major issue for predicting reactive transport in porous media systems. In this study, we introduce a framework that accurately predicts effective reaction rates directly from pore structural features by combining 3D pore-scale numerical simulations with machine learning (ML). We first perform pore-scale reactive transport simulations with fluid-solid reactions in hundreds of porous media and calculate effective reaction rates from pore-scale concentration fields. We then train a Random Forests model with 11 pore structural features and effective reaction rates to quantify the importance of structural features in determining effective reaction rates. Based on the importance information, we train artificial neural networks with varying number of features and demonstrate that effective reaction rates can be accurately predicted with only three pore structural features, which are specific surface, pore sphericity, and coordination number. Finally, global sensitivity analyses using the ML model elucidates how the three structural features affect effective reaction rates. The proposed framework enables accurate predictions of effective reaction rates directly from a few measurable pore structural features, and the framework is readily applicable to a wide range of applications involving porous media flows.

Keywords: Effective reaction rate; Machine learning; Pore-scale modeling; Porous media; Reactive transport; Fluid-solid reaction.

*Corresponding author: pkkang@umn.edu

36 Introduction

37 Predicting reactive transport in porous media is critical for a wide range of natural processes
38 as well as energy and environmental applications, including geothermal energy recovery^{1,2},
39 subsurface contaminant transport³⁻⁵, CO₂ and H₂ geological storage⁶⁻⁹, spent nuclear fuel
40 disposal¹⁰⁻¹², and water filtration^{13,14}. Reaction rate is a key input parameter to reactive
41 transport modeling, which strongly affects prediction results. A key challenge arises from the
42 fact that effective (apparent) reaction rates depend not only on intrinsic chemical properties but
43 also on pore structure and fluid flow conditions¹⁵⁻¹⁷. This is because reactive transport in
44 porous media is a strongly coupled process involving complex fluid flow, solute transport, and
45 chemical reactions. Indeed, significant discrepancies between reaction rates measured from
46 well-mixed reactors and effective reaction rates measured from column experiments and field
47 observations have been reported¹⁸⁻²¹.

48 The discrepancies between well-mixed reaction rates and effective reaction rates are known to
49 be caused by both geochemical and physical heterogeneities of porous media systems^{19,21-24}.
50 Geochemical heterogeneity originates from the variety of minerals and complexity in chemical
51 reactions²⁵⁻³¹, while physical heterogeneity is caused by the structural heterogeneity of porous
52 media, which controls fluid flow and mass transfer^{22,32-41}. In particular, pore structural
53 heterogeneity is shown to exert dominant control over fluid mixing and homogeneous reaction
54 rates⁴²⁻⁴⁴, and also shown to control porosity and permeability evolution induced by
55 heterogeneous reactions (i.e., dissolution and precipitation)^{17,34,45-50}. Yet, the quantitative
56 relationships between pore structural features (e.g., tortuosity, coordination number) that
57 characterize pore structural heterogeneity and the effective reaction rates are still elusive. This
58 limits our fundamental understanding and predictive capability of reactive transport processes
59 in porous media.

60 To uncover the quantitative relationship between pore structural features and effective reaction
61 rates in porous media, detailed representations of pore structures and reliable pore-scale
62 modeling methods are needed. Recent advances in pore-scale imaging and modeling
63 techniques enabled the accurate acquisition of pore structural information⁵¹⁻⁵⁴ and high-fidelity
64 pore-scale simulation of reactive transport^{22,32,33,38,51-55}. However, these pore-scale direct
65 numerical simulation methods are computationally demanding, which limits most studies to
66 focus on a few porous media samples. Hence, results are often specific to the studied
67 geometries, while pore geometrical complexity is enormously diverse^{56,57}.

68 Machine learning (ML) methods are considered a powerful alternative to time-consuming
69 numerical simulations while maintaining the accuracy of pore-scale direct numerical
70 simulations^{55,58-62}. Recently, deep learning frameworks based on neural networks have been
71 successfully applied to make rapid predictions of the physical properties of porous media,
72 including permeability, porosity, and specific surface area^{55,59,61,63-65}. However, ML models
73 have rarely been applied to reactive transport problems^{66,67}, and have not yet been used to
74 predict reaction rates in porous media. Furthermore, these previous ML-based investigations
75 of porous media are mostly based on a single ML algorithm. Combinations of different ML
76 algorithms, such as Random Forests (RF) and neural networks, have been used for predictive
77 diagnostics with great success in medical and bioinformatics studies, where each individual
78 algorithm is able to contribute its own strengths toward solving a specific problem^{68,69}. Yet,
79 combinations of ML models have rarely been used in reactive transport studies.

We propose an ML-based framework that characterizes the quantitative link between pore structural features and effective surface reaction rates. The pore structural features extracted from hundreds of porous media and effective reaction rates estimated from pore-scale simulations are used as the input data to train ML models. We first train an RF learning model to estimate the importance of pore structural features in determining the reaction rates. Based on this importance information, we then train an artificial neural network (ANN) model and accurately predict effective reaction rates with only three pore structural features. Finally, using the ANN model, the effects of the key pore structural features on effective reaction rates are comprehensively evaluated through global sensitivity analyses, and the physical relevance of the results is discussed.

Methods

In this section, we present the pore-scale numerical method for building the training data, and the ML methods for estimating pore structural features' importance and predicting the effective reaction rates.

Building Training Data with Pore-scale Reactive Transport Modeling

We perform a large number of pore-scale reactive transport simulations in an ensemble of 3D porous media structures obtained from an open-source database⁶⁴. Rabbani et al., (2020)⁶⁴ generated a large set of porous structures through the texture transformation and porosity manipulation of 60 measured tomographic images. Each sample has a size of 256^3 voxels, and dimensionless units are used to estimate the physical features of porous media. For each porous media, Rabbani et al., (2020)⁶⁴ quantified various pore structural features including 11 of the most common single-value pore structural features (e.g., specific surface, tortuosity, pore radius).

We apply a previously verified 3D pore-scale reactive transport model^{70–72} to calculate effective surface reaction rates in the ensemble of porous media. In this model, fluid flow, solute transport, and chemical reactions are solved. We consider incompressible fluid flow at low Reynolds numbers and solve the continuity equation and the Stokes equation⁷³,

$$\nabla \cdot \underline{v} = 0, \quad (1)$$

$$\nabla p = \mu \nabla^2 \underline{v}, \quad (2)$$

where $\underline{v}$ is the fluid velocity vector, p is the pressure, and μ is the dynamic viscosity. The governing equations of fluid flow are solved by the discrete Boltzmann equation based on the D3Q19 scheme²⁹. A no-slip boundary condition is applied via the bounce-back rule at the fluid-solid interface. Details on the applied lattice Boltzmann method can be found in Mostaghimi et al., (2016)³⁴.

The advection-diffusion equation is then solved to consider the transport of reactive solutes through pore spaces⁷⁴,

$$\frac{\partial C}{\partial t} + (\underline{v} \cdot \nabla)C = \nabla \cdot (D \nabla C), \quad (3)$$

where C is the local solute concentration, t is the time and D is the molecular diffusion coefficient. To study different flow regimes, we consider three different Péclet (Pe) numbers that cover typical flow conditions in porous media: $Pe = 0.1, 1, 10$. Pe is defined as $\frac{U_{av}L}{D}$ where

117 U_{av} is the average velocity and L is the characteristic length that is calculated via π/s_A where
 118 s_A is the specific surface area³⁴.

119 For fluid-solid reactions in the porous media, we consider a bimolecular heterogeneous
 120 reaction which can be expressed as a generic equation,

121 where A is chemical species in aqueous solutions, B is the reactant in solid phase and C is the
 122 product that can be either in aqueous phase (dissolution) or solid phase (adsorption). This
 123 equation has been reported to be adequate for describing chemical reactions in various systems
 124 and applications⁷⁵⁻⁷⁷. We implement the bimolecular heterogeneous reaction by applying
 125 irreversible first-order reaction kinetics as the boundary condition at fluid-solid interfaces via,

$$D \frac{\partial C}{\partial \underline{n}} = -k_r C, \quad (5)$$

126 where k_r represents the intrinsic reaction rate constant, and $\underline{n}$ denotes the unit normal vector
 127 to the solid surface⁴⁶. We then solve for steady-state concentration fields. The modeling
 128 framework is applicable to various types of surface reactions by using relevant reaction rate
 129 constants. In this study, we consider calcite dissolution and salt ion adsorption scenarios, where
 130 k_r is set as $1.08 \times 10^{-7} [\text{m} \cdot \text{s}^{-1}]$ ⁷⁰ and $1.0 \times 10^{-6} [\text{m} \cdot \text{s}^{-1}]$ ⁷⁸, respectively. We consider
 131 reactive transport of H^+ and Na^+ for calcite dissolution and salt ion adsorption scenarios,
 132 respectively. We use Damköhler numbers ($\text{Da} = \frac{k_r}{U_{av}}$) to describe the reaction rate relative to
 133 the mass transfer rate by advection⁷⁹, and use $\text{Da}_{II} = \text{PeDa} = \frac{k_r L}{D}$ to compare the reaction rate
 134 to the mass transfer rate by molecular diffusion^{51,80}. The diffusion coefficient of H^+ , D_{H^+} , is
 135 $1.0 \times 10^{-9} [\text{m}^2 \cdot \text{s}^{-1}]$ and D_{Na^+} is $1.3 \times 10^{-9} [\text{m}^2 \cdot \text{s}^{-1}]$ ⁸¹, and the corresponding Da_{II} values
 136 are 0.03 and 0.27. Note that Da is independent of initial concentration because the reaction
 137 follows the first-order reaction. Thus, the studied system is determined by Pe and Da (or Da_{II}).

138 From the steady-state concentration fields, we directly estimate the local reaction rate at each
 139 interfacial grid (Figure 1) and average over all interfacial grids to obtain effective reaction rates.
 140 The effective surface reaction rate can be defined as²²,

$$R_{\text{eff}} = \frac{\sum_1^N k_r (C_i - C_s)}{N} \quad (6)$$

141 where N is the total number of voxels at fluid-solid interfaces and C_i is the steady-state
 142 concentration of i -th voxel at fluid-solid interfaces. We define the normalized effective reaction
 143 rate as, $R_{\text{norm}} = \frac{R_{\text{eff}}}{k_r (C_{\text{in}} - C_s)}$, where C_{in} is the injection concentration. R_{norm} quantifies the
 144 discrepancy between the effective reaction rate and the well-mixed reaction rate for $C = C_{\text{in}}$.

145

Figure 1. The normalized local surface reaction rates shown in 2D slices (top) of two different 3D porous media (bottom).

Machine Learning Methods

The pore structures of porous media are commonly described by multiple features, such as specific surface, tortuosity, and pore radius. We use the 11 single-value pore structural features and effective reaction rates calculated from pore-scale simulations to train a two-step ML framework (Figure 2), aiming to identify the key pore structural features that control effective reaction rates. We combine RF and ANN models to quantify the importance of pore structural features and to predict effective reaction rates directly from a few key pore structural features. In comparison with neural networks, RF is less computationally expensive and can effectively estimate the importance scores of input features, i.e., RF is a useful algorithm for feature importance ranking^{82,83}. In spite of the higher computational cost for training, the ANN offers better model accuracy and performance, if the model is well optimized. Thus, we use RF to estimate feature importance, then use that importance information to optimize ANN training. Feature selection for ML models can help reduce redundant data, minimize overfitting, and improve model accuracy by removing unnecessary data. Further, ML algorithms with fewer features can be trained faster^{84,85}.

RF learning algorithm creates multiple decision trees on input data and then selects the mean predictions of each decision tree as the best solution^{82,83}. We use the data of 11 pore structural features and corresponding effective reaction rates as input to train a bagged ensemble of decision trees to estimate the importance value for each pore structural feature. The pore structural importance values are estimated by permuting out-of-bag observations among the decision trees, which are calculated as the difference between the benchmark/initial estimations and the one from the permuted predictions^{82,83}. The important hyperparameters used in the RF model training are the maximal number of decision splits (299), minimum parent size (10), minimum number of leaf node observations (1), and depth of tree (9).

ANN model applies a learning algorithm for nonlinear statistical data modeling by mimicking the way nerve cells work in the human brain, and the model is particularly efficient in implicitly estimating complex nonlinear relationships between input features and target predictions⁸⁶. We choose a single layer feed-forward neural network consisting of an input layer, hidden layer, and output layer^{87–89}. The input is the pore structural features, and effective reaction rates from pore-scale simulations are the target predictions. The hidden layer consists of 10 neurons, where multiple functions are applied for data transformation. The neurons learn about the data and then send it to the output layer. Bayesian Regularization is used as the training algorithm in ANN, which is efficient for training small-size datasets with noises. We first train the ANN with the 11 pore structural features, and then reduce the number of input features based on the importance values obtained from the RF model.

Figure 2. The schematic of the ML-based framework that combines 3D pore-scale reactive transport, Random Forests (RF), and Artificial Neural Network (ANN) models.

Results and Discussions

In this section, we first present results with an RF learning model that ranks the importance of each pore structural feature in predicting effective reaction rates. Then, we combine the RF importance ranking result with the ANN model to identify the most critical pore structural features for predicting effective reaction rates. We start with the $Pe = 0.1$ and $Da_{II} = 0.03$ case and expand to different Pe and Da regimes. Finally, we conduct global sensitivity analyses with the validated ANN model and discuss the results.

Importance Ranking of Pore Structural Features

For ML predictions, there is generally a trade-off between data preparation cost and model accuracy. ML models often show better performance with more datasets or instances for predictions, though it depends on the particular dataset^{62,90}. To test this, we train the RF learning model with pore-scale simulation results of 100, 200, 300, 400, and 500 instances (i.e., the number of porous media samples) to determine how many simulation results are needed to achieve adequate accuracy of the learning model. We use the coefficient of determination, R^2 , to measure the accuracy of the ML models⁹¹. The black square data points in Figure 3a show the R^2 values of RF models as a function of the number of instances used in training. The R^2 increases as the number of instances increases, but the increase is relatively minimal beyond 300. The R^2 of the model with 300 training instances is 0.938, which is comparable to that with 500 instances ($R^2=0.946$). Hence, we train the RF model with 300 instances and estimate the importance of 11 pore structural features.

Figure 3. (a) The R^2 for importance estimations as a function of the number of instances. Inset: the coefficient of variations of 11 pore structural features; (b) the importance ranking of 11 pore structural features estimated by RF model trained with 300 instances. The error bars represent $\pm$ standard error which is obtained from 5-fold cross-validations.

The RF successfully ranked the importance of each feature, and the results are validated with the 5-fold cross-validations. The importance rankings are shown in Figure 3b, and the importance of each feature is estimated from the increase in the prediction errors after permuting the feature in the datasets^{82,83}. The error bars are obtained from 5-fold cross-

validation, which is commonly used to test the performance of the ML model and detect overfitting⁹². The short error bars indicate that there is small uncertainty in the predicted importance values and the average values of the importance are reliable^{82,83}.

Because the importance ranking can be affected by the variability of feature values⁹⁰, we estimate the coefficient of variation of each feature, which quantifies the variability around its mean value⁹³. The inset of Figure 3a shows the coefficient of variation of each pore structural feature. It is worth noting that pore sphericity and tortuosity have lower variability than most of the other features, while they are estimated as the second and sixth most important among the 11 features. There is no noticeable correlation between the feature importance and coefficient of variation, and this indicates that the variability of feature values in the input data is large enough to evaluate the importance of pore structural features.

Predicting Effective Reaction Rates from Key Pore Structural Features

Based on the importance information, we now train the ANN model with a varying number of features to identify the key features for predicting effective reaction rates. Figure 4a shows the R^2 values for ANN predictions using 11 pore structural features. When the number of instances used for training is larger than 300, highly accurate and stable predictions are achieved, which suggests 300 instances ($R^2=0.972$) are also sufficient for establishing an accurate ANN model. The importance estimation of pore structural features by RF provides the basis for identifying key features⁹⁴⁻⁹⁶. To identify the most critical pore structural features for ANN predictions, we train the ANN using 300 instances with 11 features, and then reduce the number of input features one by one, removing the features with the lowest importance value. The inset in Figure 4a shows no significant variations in R^2 values when the ANN model is trained with three or more input features. When using one or two features for training, the R^2 values are much lower, indicating the necessity of using the three most important pore structural features for achieving accurate predictions^{64,97-99}. This implies that the other pore structural features are highly related to these three features or have limited contributions to effective reaction rates.

Specific surface, pore sphericity, and coordination number are identified as the three most important pore structural features. Specific surface area has the highest importance, which is defined as the ratio of total surface area to bulk volume⁶⁴. This result is intuitive because the specific surface is directly linked to the reactive surface area^{100,101}. Pore sphericity is a shape factor describing the smoothness of reactive surface^{97,98}, and it will affect the efficiency of mass transfer from the fluid to solid surfaces. Coordination number measures the average number of throats connected to a pore and describes the average connectivity of the pore space, which will govern the overall accessibility of reactive surface⁹⁹.

We use these three features from 300 instances as inputs to train the ANN model that predicts effective reactions rates. We use 70% of the data for training, 15% for validating, and 15% for testing the model. The testing data is independent of the training and validation data, which is applied to measure the performance of the trained ANN model. The prediction performance of the ANN model is shown in Figure 4b. The validation R^2 is 0.980, and the testing R^2 is 0.972, indicating the ANN model provides good performance in predicting effective reaction rates with only three pore structural features. The marginal discrepancy between the R^2 values from validation and testing shows the stable performance of the ML model^{90,102}.

Figure 4. (a) ANN predictions (R^2) as a function of the number of instances. Inset: the impact of the number of input features on ANN predictions using 300 instances; (b) the testing and validation performance (inset) of ANN predictions using the three pores structural features from 300 instances. The values represent normalized reaction rates.

The Effects of Pe and Da

We extend the developed framework to different flow (Pe) and reaction (Da) regimes. We first present the results under different Pe numbers. Figure 5a shows the results of importance estimation at $Pe = 0.1, 1$, and 10 with $Da_{II} = 0.03$. Specific surface, pore sphericity, and coordination number remain the three most important features, though the most important feature becomes pore sphericity at $Pe = 1$ and 10 . At higher Pe, advection is stronger, meaning pore structural features that are sensitive to flow will become more important. Indeed, the importance of pore sphericity and coordination number increases as Pe increases. The pore sphericity measures the shape of fluid-pore interfaces and determines the smoothness of the flow lines in pore space, thereby also the accessibility of the reactive surface area. Therefore, the influence of the pore shape factor on reaction increases as Pe increases. The results show that the three most important features remain the same across Pe numbers, indicating that the three features are the key predictors for effective reaction rates in typical porous media flow conditions. Figure 5b shows the test R^2 of the ANN predictions at different Pe numbers. The test R^2 values for ANN predictions at $Pe = 0.1, 1$ and 10 , are calculated as $0.972, 0.974$ and 0.980 , respectively. The R^2 values of ANN predictions are high across Pe numbers, showing stable and high-quality predictions across these different flow conditions.

We now extend the framework to a different Da number ($Da_{II} = 0.27$) with a higher intrinsic reaction constant ($1.1 \times 10^{-6} [\text{m} \cdot \text{s}^{-1}]$), which is relevant for Capacitive Deionization (CDI), an emerging desalination method⁷⁸. The transport of Na^+ and its adsorption at the solid surface is considered in the model. As shown in Figure 5c, the three most important features are the same at both $Da_{II} = 0.03$ and 0.27 with $Pe = 0.1$. The increase in reaction constant leads to a large increase in the estimated importance of the specific surface (the first green bar), which indicates that the specific surface plays a much more important role at higher reaction rates. At low flow rates ($Pe = 0.1$) but high reaction constant, the reaction rate is much larger than the advective mass transport rate, making the surface area play a more dominant role in determining the amount of fluid-solid reaction. Figure 5d shows the ANN predictions with two Da numbers. A high R^2 value of 0.976 is again achieved at $Da_{II} = 0.27$, only with the three pore

288 structural features. This result confirms that the ML framework and findings are also valid for
 289 different Da regimes.

290 **Figure 5.** (a) Importance estimation of pore structural features at three different Pe (0.1, 1, and 10) with
 291 $Da_{II} = 0.03$ (the corresponding Da values are 0.3, 0.03, 0.003) and (b) the corresponding prediction
 292 performance of ANN models using the three pore structural features from 300 instances; (c) Importance
 293 estimation of pore structural features at $Da_{II} = 0.03$ and 0.27 (the corresponding Da values are 0.3, 2.7)
 294 with $Pe = 0.1$ and (d) the corresponding prediction performance of ANN models using the three pore
 295 structural features.

296 Global Sensitivity Analysis with Machine Learning

297 We perform global sensitivity analyses using the trained ANN model to elucidate the combined
 298 effects of the key pore structural features on effective reaction rates under various flow and
 299 reaction conditions. Each row in Figure 6 shows the effects of pore structural features on
 300 normalized effective reaction rates (R_{norm}) at fixed Pe and Da . For each column, the
 301 coordination number, pore sphericity, and specific surface are fixed respectively with their
 302 average values in the datasets. This enables us to plot the combined effects of two features on

the effective reaction rates. Effective reaction rates expectedly increase with specific surface and pore sphericity, as these two features determine the area and accessibility of pore reactive surface. A larger coordination number also leads to a larger effective reaction rate because a large coordination number implies an enhanced mass transfer between pores⁶⁴.

Figure 6. The effect of specific surface, pore sphericity and coordination number on normalized effective reaction rates (R_{norm}) under different Pe and Da conditions: (a-c) at Pe = 0.1, Da = 0.3, Da_{II} = 0.03; (d-f) at Pe = 1, Da = 0.03, Da_{II} = 0.03; (g-i) at Pe = 0.1, Da = 2.7, Da_{II} = 0.27. The values of coordination number (4.6) in the left column, pore sphericity (0.76) in the middle column, and specific surface (0.13) in the right column are fixed to their average values in the dataset.

The overall effects of the three features on effective reaction rates are similar for the explored Pe and Da cases. However, the magnitude of normalized effective reaction rate is sensitive to Pe and Da. At higher Pe but low Da, the system is reaction-limited^{21,46}. In this regime, the mass transfer rate is faster than the reaction rate, such that the concentration and reaction rates in the pore space become spatially uniform, leading to the increase in R_{norm} (compare Figure 6a-b and Figure 6d-f). We confirmed that the Pe = 10 case leads to a further increase in R_{norm} , meaning less discrepancy between the well-mixed reaction rate and effective reaction rate. Figure 6g-i show the ANN predictions with Da = 2.7 at Pe = 0.1, where a large variability in effective reaction rates is estimated. This reaction regime is transport-limited due to weak advection (Pe < 1) and high reaction rate (Da > 1)^{33,46}, where the reaction rate is higher than the mass transfer rate by advection. In such regimes, uneven distributions of concentrations and

large concentration gradients emerge in the pore space, leading to discrepancies between the well-mixed reaction rate and effective reaction rate. The global sensitivity analyses elucidate the effects of the key pore structural features on effective reaction rates, which can be used as input parameters to Darcy-scale reactive transport modeling.

Conclusions

Numerous environmental applications rely on reactive transport in porous media, but the accurate estimation of reaction rates has been a major challenge, limiting the predictive capability of reactive transport models. This study established a quantitative link between pore structure features and effective surface reaction rates by combining pore-scale simulations with ML algorithms. For the first time, we identified the three key pore structural features that determine effective surface reaction rates. The three features remained as the top critical features for the explored values of Pe and Da , which cover typical flow and reaction regimes in porous media. The identified three features indeed capture the key factors that control reactive transport with heterogeneous reactions: specific surface quantifies surface area effect, pore sphericity quantifies pore shape effect, and coordination number quantifies flow/connectivity effect. We also applied this ML-based framework to perform global sensitivity analyses of the input features in determining effective reaction rates. The established ML model served as a surrogate model and enabled us to exhaustively and efficiently evaluate the effects of various system parameters (e.g., pore structures, flow rates, reaction rate constants) on effective reaction rates, which was otherwise not feasible due to the computational limitations.

The presented ML framework can be readily extended to a wide range of geological and environmental applications that involve complex coupled processes. For example, the lifetime of bentonite barriers in geologic repositories of spent nuclear fuel could be efficiently estimated by using clay structural features, temperature, chemical reaction constants, water saturation, and swelling rates as inputs to the ML model training. Further, the framework can be used to not only establish a quantitative link between input variables and target output variables but also to identify optimal values of pore structural features that can maximize the performance of porous materials. For example, by linking the pore structural features to effective corrosion rates, the framework can identify novel corrosion-resistant porous materials with optimized pore structural features that minimize corrosion. The framework can also be naturally extended to optimize other material properties such as mechanical strength and filtration efficiency. In particular, the model could identify optimal membrane properties for maximizing filtration efficiency by considering fiber diameter, hierarchical surface structure, pore size distribution, and surface area as input parameters and the effective water filtration rate as target prediction.

The proposed ML framework also provides an attractive approach for obtaining upscaled model parameters that are physically parameterized with subscale properties. In subsurface applications, the continuum model is often incapable of properly capturing pore-scale effects on Darcy-scale properties such as permeability, dispersion coefficients, and effective reaction rates. The proposed framework will enable us to establish ML-based quantitative correlations between pore-scale information and upscaled parameters. In summary, the proposed framework can not only elucidate the key parameters that control various physicochemical processes in porous media systems but also can be extended to improve model predictability and to identify optimal properties of porous materials.

References

1. Norouzi, A. M., Babaei, M., Han, W. S., Kim, K.-Y. & Niasar, V. CO₂-plume geothermal processes: A parametric study of salt precipitation influenced by capillary-driven back flow. *Chemical Engineering Journal* **425**, (2021).
2. Erfani, H., Joekar-Niasar, V. & Farajzadeh, R. Impact of Microheterogeneity on Upscaling Reactive Transport in Geothermal Energy. *ACS Earth and Space Chemistry* **3**, (2019).
3. Maher, K., Steefel, C. I., DePaolo, D. J. & Viani, B. E. The mineral dissolution rate conundrum: Insights from reactive transport modeling of U isotopes and pore fluid chemistry in marine sediments. *Geochimica et Cosmochimica Acta* **70**, (2006).
4. Lee, W. *et al.* Spatiotemporal evolution of iron and sulfate concentrations during riverbank filtration: Field observations and reactive transport modeling. *Journal of Contaminant Hydrology* **234**, (2020).
5. Zhi, W. *et al.* From Hydrometeorology to River Water Quality: Can a Deep Learning Model Predict Dissolved Oxygen at the Continental Scale? *Environmental Science & Technology* **55**, (2021).
6. Kang, Q., Lichtner, P. C., Viswanathan, H. S. & Abdel-Fattah, A. I. Pore Scale Modeling of Reactive Transport Involved in Geologic CO₂ Sequestration. *Transport in Porous Media* **82**, (2010).
7. Liu, M. & Mostaghimi, P. Pore-scale modelling of CO₂ storage in fractured coal. *International Journal of Greenhouse Gas Control* **66**, (2017).
8. Lord, A. S., Kobos, P. H. & Borns, D. J. Geologic storage of hydrogen: Scaling up to meet city transportation demands. *International Journal of Hydrogen Energy* **39**, (2014).
9. Hashemi, L., Blunt, M. & Hajibeygi, H. Pore-scale modelling and sensitivity analyses of hydrogen-brine multiphase flow in geological porous media. *Scientific Reports* **11**, (2021).
10. de Windt, L. & Spycher, N. F. Reactive Transport Modeling: A Key Performance Assessment Tool for the Geologic Disposal of Nuclear Waste. *Elements* **15**, (2019).
11. Liu, M., Kang, Q. & Xu, H. Modelling uranium dioxide corrosion under repository conditions: A pore-scale study of the chemical and thermal processes. *Corrosion Science* **167**, (2020).
12. Liu, M., Kang, Q. & Xu, H. Grain-scale study of the grain boundary effect on UO₂ fuel oxidation and fission gas release under reactor conditions. *Chemical Engineering Science* **229**, (2021).
13. Caré, S. *et al.* Modeling the Permeability Loss of Metallic Iron Water Filtration Systems. *CLEAN - Soil, Air, Water* **41**, (2013).
14. Phillip, W. A., O'Neill, B., Rodwogin, M., Hillmyer, M. A. & Cussler, E. L. Self-Assembled Block Copolymer Thin Films as Water Filtration Membranes. *ACS Applied Materials & Interfaces* **2**, (2010).
15. Ma, J., Ahkami, M., Saar, M. O. & Kong, X.-Z. Quantification of mineral accessible surface area and flow-dependent fluid-mineral reactivity at the pore scale. *Chemical Geology* **563**, (2021).
16. Ma, J., Querci, L., Hattendorf, B., Saar, M. O. & Kong, X.-Z. Toward a Spatiotemporal Understanding of Dolomite Dissolution in Sandstone by CO₂ - Enriched Brine Circulation. *Environmental Science & Technology* **53**, (2019).
17. Al-Khulaifi, Y., Lin, Q., Blunt, M. J. & Bijeljic, B. Reaction Rates in Chemically Heterogeneous Rock: Coupled Impact of Structure and Flow Properties Studied by X-ray Microtomography. *Environmental Science & Technology* **51**, (2017).

18. Maher, K., DePaolo, D. J. & Lin, J. C.-F. Rates of silicate dissolution in deep-sea sediment: In situ measurement using $^{234}\text{U}/^{238}\text{U}$ of pore fluids. *Geochimica et Cosmochimica Acta* **68**, (2004).
19. le Traon, C., Aquino, T., Bouchez, C., Maher, K. & le Borgne, T. Effective kinetics driven by dynamic concentration gradients under coupled transport and reaction. *Geochimica et Cosmochimica Acta* **306**, (2021).
20. Tong, Y. *et al.* Exploring the Utility of Compound-Specific Isotope Analysis for Assessing Ferrous Iron-Mediated Reduction of RDX in the Subsurface. *Environmental Science & Technology* **55**, 6752–6763 (2021).
21. Li, L., Steefel, C. I. & Yang, L. Scale dependence of mineral dissolution rates within single pores and fractures. *Geochimica et Cosmochimica Acta* **72**, (2008).
22. Jung, H. & Meile, C. Upscaling of microbially driven first-order reactions in heterogeneous porous media. *Journal of Contaminant Hydrology* **224**, (2019).
23. Wen, H. & Li, L. An upscaled rate law for mineral dissolution in heterogeneous media: The role of time and length scales. *Geochimica et Cosmochimica Acta* **235**, (2018).
24. Kang, P. K., Bresciani, E., An, S. & Lee, S. Potential impact of pore-scale incomplete mixing on biodegradation in aquifers: From batch experiment to field-scale modeling. *Advances in Water Resources* **123**, 1–11 (2019).
25. Atchley, A. L., Navarre-Sitchler, A. K. & Maxwell, R. M. The effects of physical and geochemical heterogeneities on hydro-geochemical transport and effective reaction rates. *Journal of Contaminant Hydrology* **165**, (2014).
26. Min, T., Gao, Y., Chen, L., Kang, Q. & Tao, W. Changes in porosity, permeability and surface area during rock dissolution: Effects of mineralogical heterogeneity. *International Journal of Heat and Mass Transfer* **103**, (2016).
27. Molins, S., Trebotich, D., Miller, G. H. & Steefel, C. I. Mineralogical and transport controls on the evolution of porous media texture using direct numerical simulation. *Water Resources Research* **53**, (2017).
28. Liu, M., Shabaninejad, M. & Mostaghimi, P. Impact of mineralogical heterogeneity on reactive transport modelling. *Computers & Geosciences* **104**, (2017).
29. Liu, M., Shabaninejad, M. & Mostaghimi, P. Predictions of permeability, surface area and average dissolution rate during reactive transport in multi-mineral rocks. *Journal of Petroleum Science and Engineering* **170**, (2018).
30. Jones, T. A. & Detwiler, R. L. Mineral Precipitation in Fractures: Using the Level-Set Method to Quantify the Role of Mineral Heterogeneity on Transport Properties. *Water Resources Research* **55**, 4186–4206 (2019).
31. Spokas, K., Peters, C. A. & Pyrak-Nolte, L. Influence of Rock Mineralogy on Reactive Fracture Evolution in Carbonate-Rich Caprocks. *Environmental Science & Technology* **52**, 10144–10152 (2018).
32. Molins, S., Trebotich, D., Steefel, C. I. & Shen, C. An investigation of the effect of pore scale flow on average geochemical reaction rates using direct numerical simulation. *Water Resources Research* **48**, (2012).
33. Deng, H., Molins, S., Trebotich, D., Steefel, C. & DePaolo, D. Pore-scale numerical investigation of the impacts of surface roughness: Upscaling of reaction rates in rough fractures. *Geochimica et Cosmochimica Acta* **239**, (2018).
34. Mostaghimi, P., Liu, M. & Arns, C. H. Numerical Simulation of Reactive Transport on Micro-CT Images. *Mathematical Geosciences* **48**, (2016).
35. Liu, M. & Mostaghimi, P. Characterisation of reactive transport in pore-scale correlated porous media. *Chemical Engineering Science* **173**, (2017).

36. Beckingham, L. E. *et al.* Evaluation of accessible mineral surface areas for improved prediction of mineral reaction rates in porous media. *Geochimica et Cosmochimica Acta* **205**, (2017).
37. Heyman, J., Lester, D. R., Turuban, R., Méheust, Y. & le Borgne, T. Stretching and folding sustain microscale chemical gradients in porous media. *Proceedings of the National Academy of Sciences* **117**, (2020).
38. Yang, Y. *et al.* Dynamic Pore-Scale Dissolution by CO₂-Saturated Brine in Carbonates: Impact of Homogeneous Versus Fractured Versus Vuggy Pore Structure. *Water Resources Research* **56**, (2020).
39. Yoon, S. & Kang, P. K. Roughness, inertia, and diffusion effects on anomalous transport in rough channel flows. *Physical Review Fluids* **6**, 014502 (2021).
40. Kanavas, Z., Pérez-Reche, F. J., Arns, F. & Morales, V. L. Flow Path Resistance in Heterogeneous Porous Media Recast into a Graph-Theory Problem. *Transport in Porous Media* (2021) doi:10.1007/s11242-021-01671-6.
41. de Anna, P., Quaipe, B., Biros, G. & Juanes, R. Prediction of the low-velocity distribution from the pore structure in simple porous media. *Physical Review Fluids* **2**, 124103 (2017).
42. Boon, M., Bijeljic, B. & Krevor, S. Observations of the impact of rock heterogeneity on solute spreading and mixing. *Water Resources Research* **53**, (2017).
43. Dentz, M., le Borgne, T., Englert, A. & Bijeljic, B. Mixing, spreading and reaction in heterogeneous media: A brief review. *Journal of Contaminant Hydrology* **120–121**, (2011).
44. Alhashmi, Z., Blunt, M. J. & Bijeljic, B. The Impact of Pore Structure Heterogeneity, Transport, and Reaction Conditions on Fluid–Fluid Reaction Rate Studied on Images of Pore Space. *Transport in Porous Media* **115**, (2016).
45. Liu, M. & Mostaghimi, P. Pore-scale simulation of dissolution-induced variations in rock mechanical properties. *International Journal of Heat and Mass Transfer* **111**, (2017).
46. Kang, Q., Chen, L., Valocchi, A. J. & Viswanathan, H. S. Pore-scale study of dissolution-induced changes in permeability and porosity of porous media. *Journal of Hydrology* **517**, (2014).
47. Chen, L., Kang, Q., Viswanathan, H. S. & Tao, W.-Q. Pore-scale study of dissolution-induced changes in hydrologic properties of rocks with binary minerals. *Water Resources Research* **50**, (2014).
48. Jiménez-Martínez, J. *et al.* Homogenization of Dissolution and Enhanced Precipitation Induced by Bubbles in Multiphase Flow Systems. *Geophysical Research Letters* **47**, (2020).
49. Molins, S. *et al.* Simulation of mineral dissolution at the pore scale with evolving fluid-solid interfaces: review of approaches and benchmark problem set. *Computational Geosciences* **25**, 1285–1318 (2021).
50. Starchenko, V. & Ladd, A. J. C. The Development of Wormholes in Laboratory-Scale Fractures: Perspectives From Three-Dimensional Simulations. *Water Resources Research* **54**, 7946–7959 (2018).
51. Menke, H. P., Bijeljic, B., Andrew, M. G. & Blunt, M. J. Dynamic Three-Dimensional Pore-Scale Imaging of Reaction in a Carbonate at Reservoir Conditions. *Environmental Science & Technology* **49**, 4407–4414 (2015).
52. Wildenschild, D. & Sheppard, A. P. X-ray imaging and analysis techniques for quantifying pore-scale structure and processes in subsurface porous medium systems. *Advances in Water Resources* **51**, (2013).

53. Ghosh, S., Ohashi, H., Tabata, H., Hashimasa, Y. & Yamaguchi, T. Microstructural pore analysis of the catalyst layer in a polymer electrolyte membrane fuel cell: A combination of resin pore-filling and FIB/SEM. *International Journal of Hydrogen Energy* **40**, 15663–15671 (2015).
54. Noiriel, C. & Soullaine, C. Pore-Scale Imaging and Modelling of Reactive Flow in Evolving Porous Media: Tracking the Dynamics of the Fluid–Rock Interface. *Transport in Porous Media* **140**, 181–213 (2021).
55. Kamrava, S., Tahmasebi, P. & Sahimi, M. Linking Morphology of Porous Media to Their Macroscopic Permeability by Deep Learning. *Transport in Porous Media* **131**, (2020).
56. Chung, T., Wang, Y. da, Armstrong, R. T. & Mostaghimi, P. Minimising the impact of sub-resolution features on fluid flow simulation in porous media. *Journal of Petroleum Science and Engineering* **207**, (2021).
57. Wang, Y. da, Chung, T., Armstrong, R. T., McClure, J. E. & Mostaghimi, P. Computations of permeability of large rock images by dual grid domain decomposition. *Advances in Water Resources* **126**, (2019).
58. Russell, S. & Norvig, P. *Artificial Intelligence: A Modern Approach*. (Prentice Hall, 2020).
59. Wang, Y. da, Blunt, M. J., Armstrong, R. T. & Mostaghimi, P. Deep learning in pore scale imaging and modeling. *Earth-Science Reviews* **215**, (2021).
60. Rabbani, A. *et al.* Review of Data Science Trends and Issues in Porous Media Research With a Focus on Image-Based Techniques. *Water Resources Research* **57**, (2021).
61. Menke, H. P., Maes, J. & Geiger, S. Upscaling the porosity–permeability relationship of a microporous carbonate for Darcy-scale flow with machine learning. *Scientific Reports* **11**, (2021).
62. Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **349**, (2015).
63. Alqahtani, N., Alzubaidi, F., Armstrong, R. T., Swietojanski, P. & Mostaghimi, P. Machine learning for predicting properties of porous media from 2d X-ray images. *Journal of Petroleum Science and Engineering* **184**, (2020).
64. Rabbani, A., Babaei, M., Shams, R., Wang, Y. da & Chung, T. DeePore: A deep learning workflow for rapid and comprehensive characterization of porous materials. *Advances in Water Resources* **146**, (2020).
65. Chung, T., Wang, Y. da, Armstrong, R. T. & Mostaghimi, P. Voxel agglomeration for accelerated estimation of permeability from micro-CT images. *Journal of Petroleum Science and Engineering* **184**, (2020).
66. Leal, A. M. M., Kvas, S., Kulik, D. A. & Saar, M. O. Accelerating Reactive Transport Modeling: On-Demand Machine Learning Algorithm for Chemical Equilibrium Calculations. *Transport in Porous Media* **133**, (2020).
67. Guérillot, D. & Bruyelle, J. Geochemical equilibrium determination using an artificial neural network in compositional reservoir flow simulation. *Computational Geosciences* **24**, (2020).
68. Maji, D., Santara, A., Ghosh, S., Sheet, D. & Mitra, P. Deep neural network and random forest hybrid architecture for learning to detect retinal vessels in fundus images. in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2015). doi:10.1109/EMBC.2015.7319030.

69. Kong, Y. & Yu, T. A Deep Neural Network Model using Random Forest to Extract Feature Representation for Gene Expression Data Classification. *Scientific Reports* **8**, (2018).
70. Liu, M. & Mostaghimi, P. High-resolution pore-scale simulation of dissolution in porous media. *Chemical Engineering Science* **161**, (2017).
71. Liu, M. & Mostaghimi, P. Numerical simulation of fluid-fluid-solid reactions in porous media. *International Journal of Heat and Mass Transfer* **120**, (2018).
72. Liu, M. & Mostaghimi, P. Reactive transport modelling in dual porosity media. *Chemical Engineering Science* **190**, (2018).
73. Sahimi, M. *Flow and Transport in Porous Media and Fractured Rock: From Classical Methods to Modern Approaches*. (John Wiley & Sons, 2011).
74. Adler, P. *Porous media: geometry and transports*. (Elsevier, 2013).
75. Coppens, M.-O. & Froment, G. F. Diffusion and reaction in a fractal catalyst pore—II. Diffusion and first-order reaction. *Chemical Engineering Science* **50**, 1027–1039 (1995).
76. Pereira Nunes, J. P., Bijeljic, B. & Blunt, M. J. Pore-space structure and average dissolution rates: A simulation study. *Water Resources Research* **52**, 7198–7212 (2016).
77. Ryan, E. M., Tartakovsky, A. M. & Amon, C. Pore-scale modeling of competitive adsorption in porous media. *Journal of Contaminant Hydrology* **120–121**, 56–78 (2011).
78. Li, G., Cai, W., Zhao, R. & Hao, L. Electrosorptive removal of salt ions from water by membrane capacitive deionization (MCDI): characterization, adsorption equilibrium, and kinetics. *Environmental Science and Pollution Research* **26**, (2019).
79. Lasaga, A. C. Chemical kinetics of water-rock interactions. *Journal of Geophysical Research: Solid Earth* **89**, 4009–4025 (1984).
80. Kang, Q., Zhang, D. & Chen, S. Simulation of dissolution and precipitation in porous media. *Journal of Geophysical Research: Solid Earth* **108**, (2003).
81. Lide, D. R. *CRC handbook of chemistry and physics*. vol. 85 (CRC press, 2004).
82. Kwon, B., Ejaz, F. & Hwang, L. K. Machine learning for heat transfer correlations. *International Communications in Heat and Mass Transfer* **116**, (2020).
83. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
84. Liu, H. & Motoda, H. *Computational methods of feature selection*. (CRC Press, 2007).
85. Guyon, I., Gunn, S., Nikravesh, M. & Zadeh, L. A. *Feature Extraction: Foundations and Applications*. vol. 207 (Springer, 2008).
86. Dreyfus, S. E. Artificial neural networks, back propagation, and the Kelley-Bryson gradient procedure. *Journal of Guidance, Control, and Dynamics* **13**, (1990).
87. Bebis, G. & Georgiopoulos, M. Feed-forward neural networks. *IEEE Potentials* **13**, (1994).
88. Lei, X., Liao, X., Chen, F. & Huang, T. Two-layer tree-connected feed-forward neural network model for neural cryptography. *Physical Review E* **87**, 32811 (2013).
89. Nguyen-Truong, H. T. & Le, H. M. An implementation of the Levenberg–Marquardt algorithm for simultaneous-energy-gradient fitting using two-layer feed-forward neural networks. *Chemical Physics Letters* **629**, (2015).
90. Mohri, M., Rostamizadeh, A. & Talwalkar, A. *Foundations of machine learning*. (MIT press, 2018).
91. di Bucchianico, A. Coefficient of Determination (R^2). in *Encyclopedia of Statistics in Quality and Reliability* (John Wiley & Sons, Ltd, 2008). doi:10.1002/9780470061572.eqr173.

92. Rodriguez, J. D., Perez, A. & Lozano, J. A. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **32**, (2010).
93. Brown, C. E. Coefficient of Variation. in *Applied Multivariate Statistics in Geohydrology and Related Sciences* (Springer Berlin Heidelberg, 1998). doi:10.1007/978-3-642-80328-4_13.
94. Cunningham, P. Dimension Reduction. in *Machine Learning Techniques for Multimedia* (Springer Berlin Heidelberg, 2008). doi:10.1007/978-3-540-75171-7_4.
95. Panthong, R. & Srivihok, A. Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm. *Procedia Computer Science* **72**, (2015).
96. Chizi, B. & Maimon, O. Dimension Reduction and Feature Selection. in *Data Mining and Knowledge Discovery Handbook* (Springer US, 2009). doi:10.1007/978-0-387-09823-4_5.
97. Schmitt, M. , H. M. , M. C. , and F. C. P. Classification and quantification of pore shapes in sandstone reservoir rocks with 3-D X-ray micro-computed tomography. *Solid Earth* **7**, 285–300 (2016).
98. Kong, L., Ostadhassan, M., Li, C. & Tamimi, N. Pore characterization of 3D-printed gypsum rocks: a comprehensive approach. *Journal of Materials Science* **53**, (2018).
99. Blunt, M. J. *et al.* Pore-scale imaging and modelling. *Advances in Water Resources* **51**, (2013).
100. Gouze, P. & Luquot, L. X-ray microtomography characterization of porosity, permeability and reactive surface changes during dissolution. *Journal of Contaminant Hydrology* **120–121**, 45–55 (2011).
101. Noiriel, C. *et al.* Changes in reactive surface area during limestone dissolution: An experimental and modelling study. *Chemical Geology* **265**, 160–170 (2009).
102. Xie, X. *et al.* Testing and validating machine learning classifiers by metamorphic testing. *Journal of Systems and Software* **84**, (2011).

Acknowledgements

Peter K. Kang acknowledges the support by the National Science Foundation under Grant No. EAR-2046015. Beomjin Kwon acknowledges the support by the National Science Foundation under Grant No. CBET-2053413.

Author contributions

M.L. conducted the pore-scale simulations, trained the ML models, analyzed the results, and wrote the first draft of the manuscript. B.K. contributed to the ML methods and application, analyzed the results, and edited the manuscript. P.K.K. was the principal investigator of the research, designed the research, analyzed the results, and edited the manuscript.

Data Availability Statement

The training data and trained ML models are all made available open access at https://drive.google.com/drive/u/2/folders/17nTPOjOVslivzZG8u0_-l4gibCV3vzHX.

Competing interests

The authors declare no competing interests.

Correspondence and requests for materials should be addressed to P.K.K.