



OPEN

# Machine learning to predict effective reaction rates in 3D porous media from pore structural features

Min Liu<sup>1,2</sup>, Beomjin Kwon<sup>3</sup> & Peter K. Kang<sup>1,2</sup>✉

Large discrepancies between well-mixed reaction rates and effective reaction 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.

Predicting reactive transport in porous media is critical for a wide range of natural processes as well as energy and environmental applications, including geothermal energy recovery<sup>1,2</sup>, subsurface contaminant transport<sup>3–5</sup>, CO<sub>2</sub> and H<sub>2</sub> geological storage<sup>6–9</sup>, spent nuclear fuel disposal<sup>10–12</sup>, and water filtration<sup>13,14</sup>. Reaction rate is a key input parameter to reactive transport modeling, which strongly affects prediction results. A key challenge arises from the fact that effective (apparent) reaction rates depend not only on intrinsic chemical properties but also on pore structure and fluid flow conditions<sup>15–17</sup>. This is because reactive transport in porous media is a strongly coupled process involving complex fluid flow, solute transport, and chemical reactions. Indeed, significant discrepancies between reaction rates measured from well-mixed reactors and effective reaction rates measured from column experiments and field observations have been reported<sup>18–21</sup>.

The discrepancies between well-mixed reaction rates and effective reaction rates are known to be caused by both geochemical and physical heterogeneities of porous media systems<sup>19,21–24</sup>. Geochemical heterogeneity originates from the variety of minerals and complexity in chemical reactions<sup>25–31</sup>, while physical heterogeneity is caused by the structural heterogeneity of porous media, which controls fluid flow and mass transfer<sup>22,32–41</sup>. In particular, pore structural heterogeneity is shown to exert dominant control over fluid mixing and homogeneous reaction rates<sup>42–44</sup>, and also shown to control porosity and permeability evolution induced by heterogeneous reactions (i.e., dissolution and precipitation)<sup>17,34,45–50</sup>. Yet, the quantitative relationships between pore structural features (e.g., tortuosity, coordination number) that characterize pore structural heterogeneity and the effective reaction rates are still elusive. This limits our fundamental understanding and predictive capability of reactive transport processes in porous media.

To uncover the quantitative relationship between pore structural features and effective reaction rates in porous media, detailed representations of pore structures and reliable pore-scale modeling methods are needed. Recent advances in pore-scale imaging and modeling techniques enabled the accurate acquisition of pore structural information<sup>51–54</sup> and high-fidelity pore-scale simulation of reactive transport<sup>22,32,33,38,51–55</sup>. However, these

<sup>1</sup>Department of Earth and Environmental Sciences, University of Minnesota, Minneapolis, USA. <sup>2</sup>Saint Anthony Falls Laboratory, University of Minnesota, Minneapolis, USA. <sup>3</sup>School for Engineering of Matter, Transport and Energy, Arizona State University, Phoenix, USA. ✉email: pkkang@umn.edu

pore-scale direct numerical simulation methods are computationally demanding, which limits most studies to focus on a few porous media samples. Hence, results are often specific to the studied geometries, while pore geometrical complexity is enormously diverse<sup>56,57</sup>.

Machine learning (ML) methods are considered as a powerful alternative to time-consuming numerical simulations while maintaining the accuracy of pore-scale direct numerical simulations<sup>55,58–62</sup>. Recently, deep learning frameworks based on neural networks have been successfully applied to make rapid predictions of the physical properties of porous media, including permeability, porosity, and specific surface area<sup>55,59,61,63–65</sup>. However, ML models have rarely been applied to reactive transport problems<sup>66,67</sup>, and have not yet been used to predict reaction rates in porous media. Furthermore, these previous ML-based investigations of porous media are mostly based on a single ML algorithm. Combinations of different ML algorithms, such as Random Forests (RF) and neural networks, have been used for predictive diagnostics with great success in medical and bioinformatics studies, where each individual algorithm is able to contribute its own strengths toward solving a specific problem<sup>68,69</sup>. Yet, 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. We demonstrate the developed framework in different flow (Pe) and reaction (Da) regimes. 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 database<sup>64</sup>. Rabbani et al.<sup>64</sup> 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.<sup>64</sup> quantified various pore structural features including 11 of the most common single-value pore structural features. They are specific surface, pore sphericity, coordination number, throat radius, pore radius, tortuosity, pore-throat ratio, grain radius, pore density, grain sphericity, and throat length.

We apply a previously verified 3D pore-scale reactive transport model<sup>70–72</sup> 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 equation<sup>73</sup>,

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

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

where  $\underline{v}$  is the fluid velocity vector,  $p$  is the pressure, and  $\mu$  is the dynamic viscosity. The governing equations of fluid flow are solved by the discrete Boltzmann equation based on the D3Q19 scheme<sup>29</sup>. 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.<sup>34</sup>.

The advection–diffusion equation is then solved to consider the transport of reactive solutes through pore spaces<sup>74</sup>,

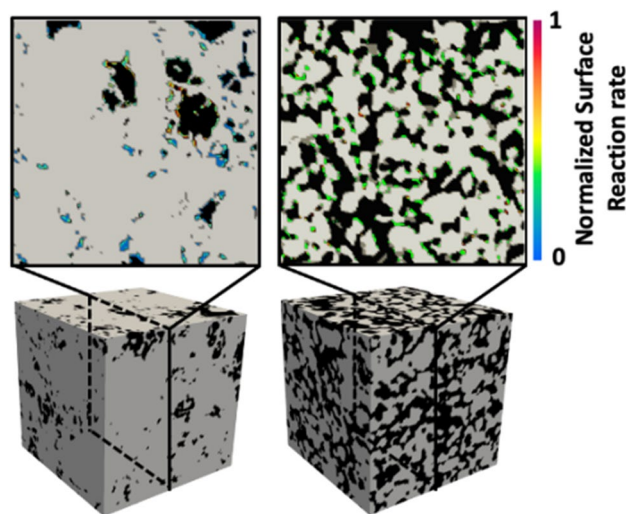
$$\frac{\partial C}{\partial t} + (\underline{v} \cdot \nabla)C = \nabla \cdot (D\nabla C), \quad (3)$$

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  $U_{av}$  is the average velocity and  $L$  is the characteristic length that is calculated via  $\pi/s_A$  where  $s_A$  is the specific surface area<sup>34</sup>.

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



where  $A$  is chemical species in aqueous solutions,  $B$  is the reactant in solid phase and  $C$  is the product that can be either in aqueous phase (dissolution) or solid phase (adsorption). This equation has been reported to be adequate for describing chemical reactions in various systems and applications<sup>75–77</sup>. We implement the bimolecular heterogeneous reaction by applying irreversible first-order reaction kinetics as the boundary condition at fluid–solid interfaces via,



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

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

where  $k_r$  represents the intrinsic reaction rate constant, and  $n$  denotes the unit normal vector to the solid surface<sup>46</sup>. We then solve for steady-state concentration fields. The modeling framework is applicable to various types of surface reactions by using relevant reaction rate constants. In this study, we consider calcite dissolution and salt ion adsorption scenarios, where  $k_r$  is set as  $1.08 \times 10^{-7} [\text{m s}^{-1}]$ <sup>51,70</sup> and  $1.0 \times 10^{-6} [\text{m s}^{-1}]$ <sup>78</sup>, respectively. The dissolution rate constant for calcite is measured in  $\text{CO}_2$ -saturated water in batch reactors at 323 K and at 10 MPa<sup>51</sup>, and the salt ion adsorption rate constant is measured in a carbon membrane capacitive deionization cell with a potential of 1.2 V at pH 7.0<sup>78</sup>. We consider reactive transport of  $\text{H}^+$  and  $\text{Na}^+$  for calcite dissolution and salt ion adsorption scenarios, respectively. We use Damköhler numbers ( $\text{Da} = \frac{k_r L}{U_{av}}$ ) to describe the reaction rate relative to the mass transfer rate by advection<sup>79</sup>, and use  $\text{Da}_{II} = \text{Pe Da} = \frac{k_r L}{D}$  to compare the reaction rate to the mass transfer rate by molecular diffusion<sup>51,80</sup>. The diffusion coefficient of  $\text{H}^+$ ,  $D_{\text{H}^+}$ , is  $1.0 \times 10^{-9} [\text{m}^2 \text{s}^{-1}]$  and  $D_{\text{Na}^+}$  is  $1.3 \times 10^{-9} [\text{m}^2 \text{s}^{-1}]$ <sup>81</sup>, and the corresponding  $\text{Da}_{II}$  values are 0.03 and 0.27. Note that  $\text{Da}$  is independent of initial concentration because the reaction follows the first-order reaction. Thus, the studied system is determined by  $\text{Pe}$  and  $\text{Da}$  (or  $\text{Da}_{II}$ ).

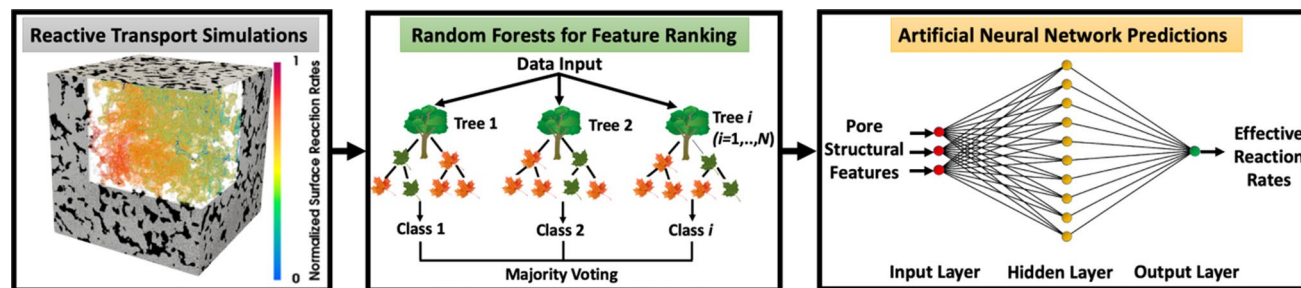
From the steady-state concentration fields, we directly estimate the local reaction rate at each interfacial grid (Fig. 1) and average over all interfacial grids to obtain effective reaction rates. The effective surface reaction rate can be defined as<sup>22</sup>,

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

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

**Machine learning methods.** The pore structures of porous media can be characterized 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 (Fig. 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<sup>82,83</sup>. 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<sup>84,85</sup>.

RF learning algorithm creates multiple decision trees on input data and then selects the mean predictions of each decision tree as the best solution<sup>82,83</sup>. 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



**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.

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<sup>82,83</sup>. 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). The hyperparameters are optimized to maximize  $R^2$  values and to ensure enough splits and tree depth.

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 predictions<sup>86</sup>. We choose a single layer feed-forward neural network consisting of an input layer, hidden layer, and output layer<sup>87–89</sup>. 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.

## Results and discussion

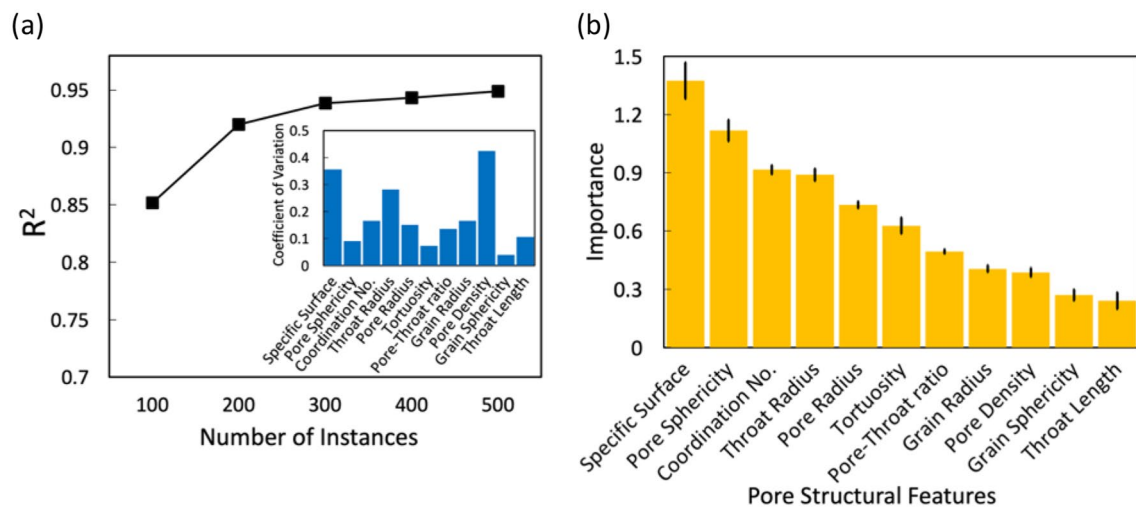
In this section, we first present results with an RF learning model that ranks the importance of each pore structural feature in predicting effective reaction rates. Then, we combine the RF importance ranking result with the ANN model to identify the most critical pore structural features for predicting 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<sup>62,90</sup>. 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 models<sup>91</sup>. The black square data points in Fig. 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.

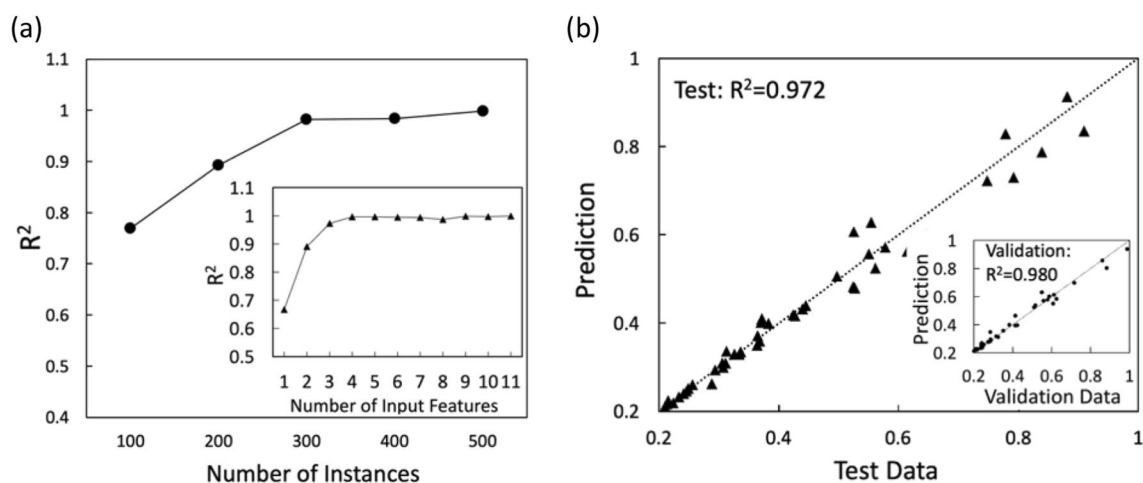
The RF successfully ranked the importance of each feature, and the results are validated with the fivefold cross-validations. The importance rankings are shown in Fig. 3b, and the importance of each feature is estimated from the increase in the prediction errors after permuting the feature in the datasets<sup>82,83</sup>. The error bars are obtained from fivefold cross-validation, which is commonly used to test the performance of the ML model and detect overfitting<sup>92</sup>. The short error bars indicate that there is small uncertainty in the predicted importance values and the average values of the importance are reliable<sup>82,83</sup>.

Because the importance ranking can be affected by the variability of feature values<sup>90</sup>, we estimate the coefficient of variation of each feature, which quantifies the variability around its mean value<sup>93</sup>. The inset of Fig. 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 pre-



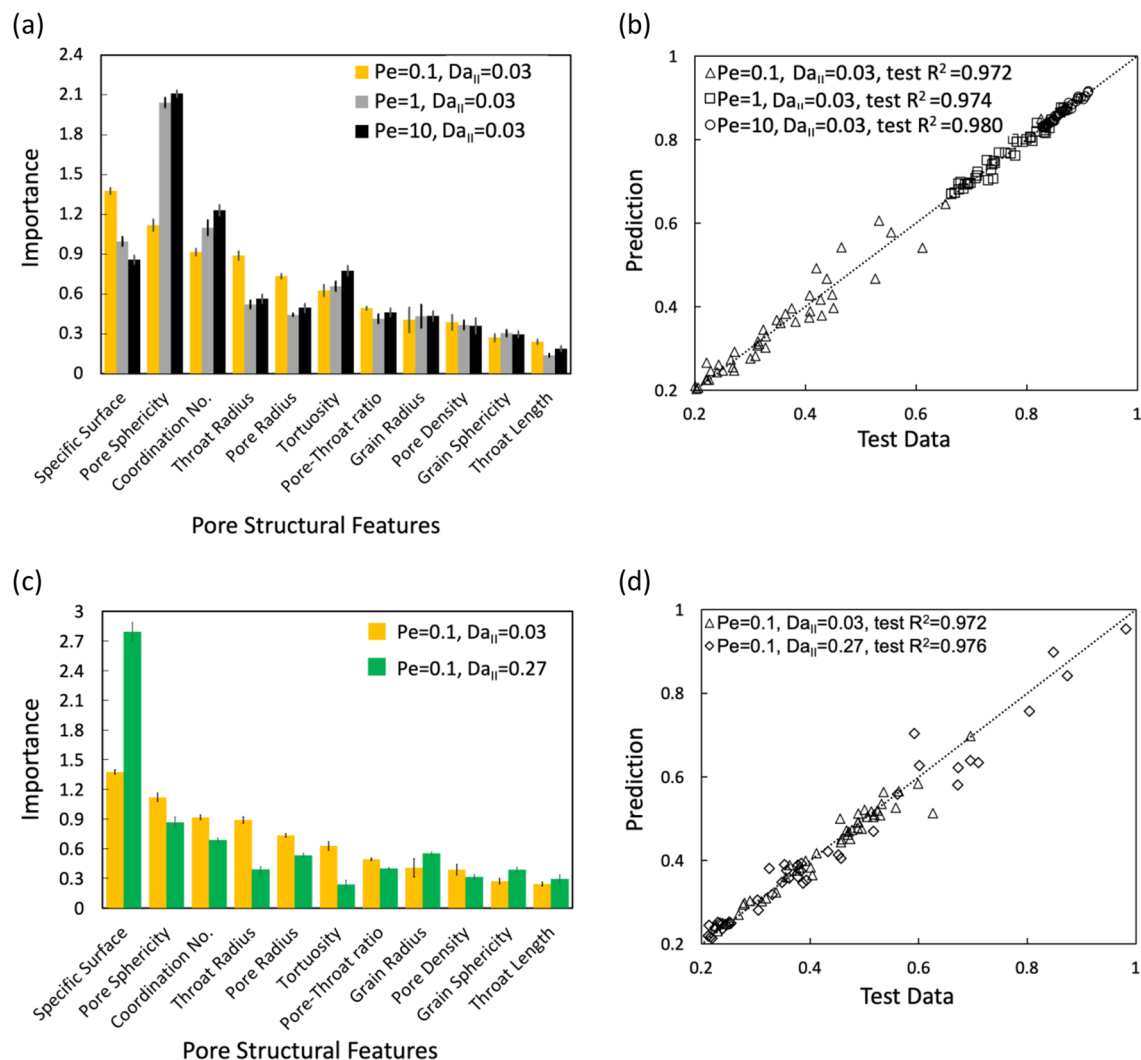
**Figure 3.** (a) The  $R^2$  for importance estimations as a function of the number of instances. Inset: the coefficient of variations of 11 pore structural features; (b) the importance ranking of 11 pore structural features estimated by RF model trained with 300 instances. The error bars represent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



**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. Features are removed one by one starting from the least important feature; (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.

dictions 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 features<sup>94–96</sup>. 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 importance ranking from the most important feature to the least important feature are the following: specific surface, pore sphericity, coordination number, throat radius, pore radius, tortuosity, pore-throat ratio, grain radius, pore density, grain sphericity, throat length. The inset in Fig. 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<sup>64,97–99</sup>. This implies that the other pore structural features are highly related to these three features or have limited contributions to effective reaction rates, which can be confirmed from the pair-wise correlations between the 11 features (see the Supplementary Document). For example, throat radius shows a high correlation with pore sphericity, and tortuosity shows a high correlation with coordination number.

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 volume<sup>64</sup>. This result is intuitive because the specific surface is directly linked to the reactive surface

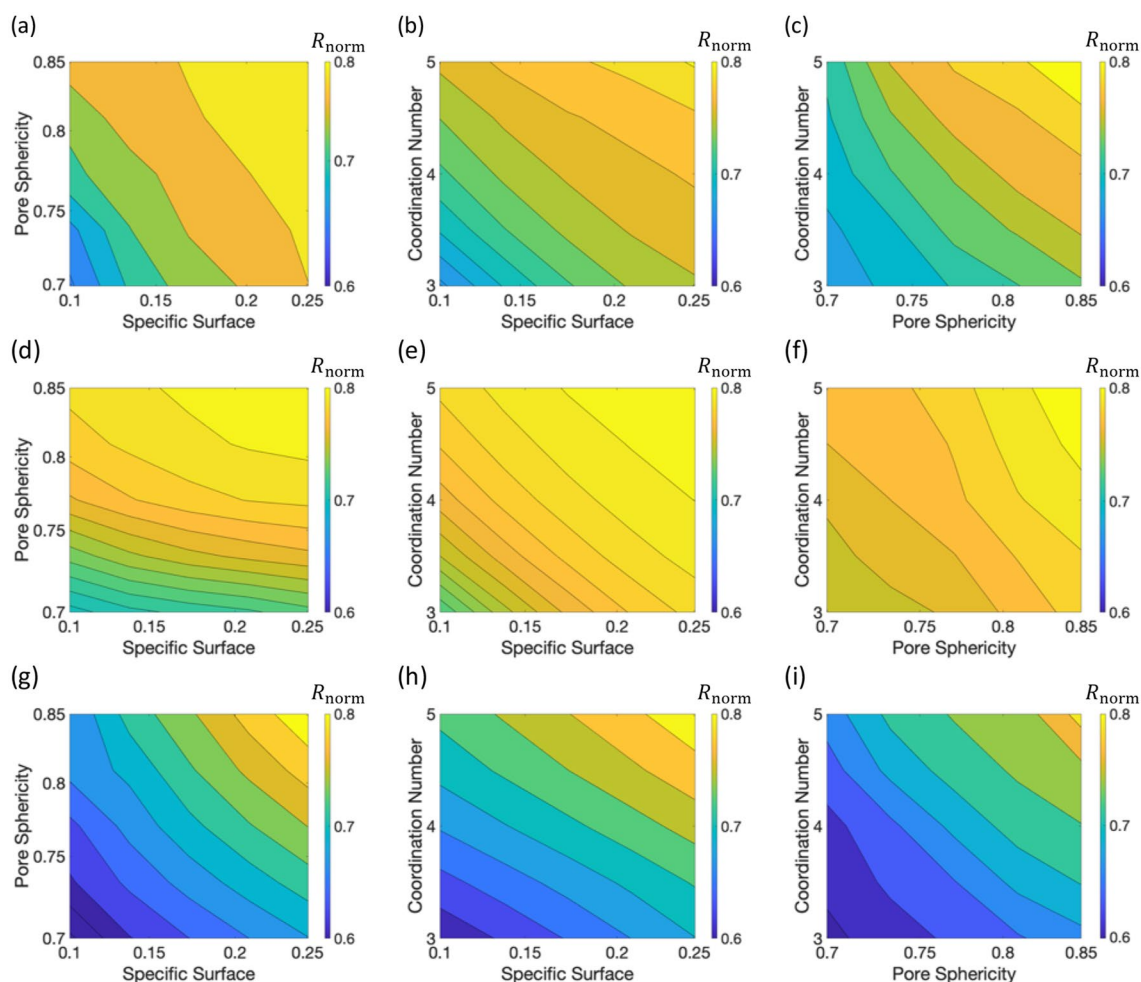


**Figure 5.** (a) Importance estimation of pore structural features at three different Pe (0.1, 1, and 10) with  $Da_{II}=0.03$  (the corresponding Da values are 0.3, 0.03, 0.003) and (b) the corresponding prediction performance of ANN models using the three pore structural features from 300 instances;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

area<sup>100,101</sup>. Pore sphericity is a shape factor describing the smoothness of reactive surface<sup>97,98</sup>, 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 surface<sup>99</sup>.

We use these three features from 300 instances as inputs to train the ANN model that predicts effective reaction 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 Fig. 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<sup>90,102</sup>.

**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



**Figure 6.** The effect of specific surface, pore sphericity and coordination number on normalized effective reaction rates ( $R_{\text{norm}}$ ) under different Pe and Da conditions: (a–c) at Pe = 0.1, Da = 0.3,  $Da_{\text{II}} = 0.03$ ; (d–f) at Pe = 1, Da = 0.03,  $Da_{\text{II}} = 0.03$ ; (g–i) at Pe = 0.1, Da = 2.7,  $Da_{\text{II}} = 0.27$ . 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 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 ( $Da_{\text{II}} = 0.27$ ) with a higher intrinsic reaction constant ( $1.1 \times 10^{-6} [\text{ms}^{-1}]$ ), which is relevant for Capacitive Deionization (CDI), an emerging desalination method<sup>78</sup>. The transport of  $\text{Na}^+$  and its adsorption at the solid surface is considered in the model. As shown in Fig. 5c, the three most important features are the same at both  $Da_{\text{II}} = 0.03$  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  $Da_{\text{II}} = 0.27$ , only with the three pore structural features. This result confirms that the ML framework and findings are also valid for different Da regimes.

**Global sensitivity analysis with machine learning.** We perform global sensitivity analyses using the trained ANN model to elucidate the combined effects of the key pore structural features on effective reaction rates under various flow and reaction conditions. Each row in Fig. 6 shows the effects of pore structural features on normalized effective reaction rates ( $R_{\text{norm}}$ ) at fixed Pe and Da. For each column, the coordination number, pore sphericity, and specific surface are fixed respectively with their 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 pores<sup>64</sup>.

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<sup>21,46</sup>. 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  $R_{\text{norm}}$  (compare Fig. 6a,b and d–f). We confirmed that the Pe = 10 case leads to a further increase in  $R_{\text{norm}}$ , 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)<sup>33,46</sup>, 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, and the effective reaction rates could be used as input parameters to Darcy-scale reactive transport modeling.

If solid phase alteration is considered, uniform or transitional (between uniform and wormholing) dissolution patterns are most likely to be observed due to the low Da numbers<sup>35,103</sup>. However, it is known that the dissolution patterns depend not only on Pe and Da but also on pore structural heterogeneity<sup>38,46,104</sup>. The ML framework proposed in this study could be extended to predict dissolution regimes from Pe, Da, and pore structural features. However, reactive flow simulations with solid alterations, which is computationally very expensive, should be performed to obtain accurate dissolution patterns.

## 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<sup>105,106</sup>. 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. Extending the applicability of the proposed framework to wider ranges of Pe and Da will be an important next step. Also, with a larger dataset with wider Pe and Da values, one may be able to develop a more generic ML model that includes Pe and Da as input parameters.

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 a future study, the effects of porous media sample size should be further investigated. Studying the sample size effect will require substantial computational resources, but it is an important step for achieving upscaling. 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.

## Data availability

The training data and trained ML models are all made available open access at [https://drive.google.com/drive/u/2/folders/17nTPOjOVslivzZG8u0\\_-l4gibCV3vzHX](https://drive.google.com/drive/u/2/folders/17nTPOjOVslivzZG8u0_-l4gibCV3vzHX).

Received: 21 January 2022; Accepted: 24 March 2022

Published online: 31 March 2022



## References

- Norouzi, A. M., Babaei, M., Han, W. S., Kim, K.-Y. & Niasar, V. CO<sub>2</sub>-plume geothermal processes: A parametric study of salt precipitation influenced by capillary-driven back flow. *Chem. Eng. J.* **425**, 130031 (2021).
- Erfani, H., Joekar-Niasar, V. & Farajzadeh, R. Impact of microheterogeneity on upscaling reactive transport in geothermal energy. *ACS Earth Space Chem.* **3**, 2045 (2019).
- 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. *Geochim. et Cosmochim. Acta* **70**, 337 (2006).
- Lee, W. *et al.* Spatiotemporal evolution of iron and sulfate concentrations during riverbank filtration: Field observations and reactive transport modeling. *J. Contam. Hydrol.* **234**, 103697 (2020).
- Zhi, W. *et al.* From hydrometeorology to river water quality: Can a deep learning model predict dissolved oxygen at the continental scale? *Environ. Sci. Technol.* **55**, 2357 (2021).
- Kang, Q., Lichtner, P. C., Viswanathan, H. S. & Abdel-Fattah, A. I. Pore scale modeling of reactive transport involved in geologic CO<sub>2</sub> sequestration. *Transp. Porous Media* **82**, 197 (2010).
- Liu, M. & Mostaghimi, P. Pore-scale modelling of CO<sub>2</sub> storage in fractured coal. *Int. J. Greenhouse Gas Control* **66**, 246 (2017).
- Lord, A. S., Kobos, P. H. & Borns, D. J. Geologic storage of hydrogen: Scaling up to meet city transportation demands. *Int. J. Hydrogen Energy* **39**, 15570 (2014).
- Hashemi, L., Blunt, M. & Hajibeygi, H. Pore-scale modelling and sensitivity analyses of hydrogen-brine multiphase flow in geological porous media. *Sci. Rep.* <https://doi.org/10.1038/s41598-021-87490-7> (2021).
- de Windt, L. & Spycher, N. F. Reactive transport modeling: A key performance assessment tool for the geologic disposal of nuclear waste. *Elements* **15**, 99 (2019).
- Liu, M., Kang, Q. & Xu, H. Modelling uranium dioxide corrosion under repository conditions: A pore-scale study of the chemical and thermal processes. *Corros. Sci.* **167**, 108530 (2020).
- Liu, M., Kang, Q. & Xu, H. Grain-scale study of the grain boundary effect on UO<sub>2</sub> fuel oxidation and fission gas release under reactor conditions. *Chem. Eng. Sci.* **229**, 116025 (2021).
- Caré, S. *et al.* Modeling the permeability loss of metallic iron water filtration systems. *Clean: Soil, Air, Water* **41**, 275 (2013).
- 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 Appl. Mater. Interfaces* **2**, 847 (2010).
- 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. *Chem. Geol.* **563**, 120042 (2021).
- Ma, J., Querci, L., Hattendorf, B., Saar, M. O. & Kong, X.-Z. Toward a spatiotemporal understanding of dolomite dissolution in sandstone by CO<sub>2</sub>-enriched brine circulation. *Environ. Sci. Technol.* **53**, 12458 (2019).
- 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. *Environ. Sci. Technol.* **51**, 4108 (2017).
- Maher, K., DePaolo, D. J. & Lin, J.C.-F. Rates of silicate dissolution in deep-sea sediment: In situ measurement using <sup>234</sup>U/<sup>238</sup>U of pore fluids. *Geochim. et Cosmochim. Acta* **68**, 4629 (2004).
- le Traon, C., Aquino, T., Bouchez, C., Maher, K. & le Borgne, T. Effective kinetics driven by dynamic concentration gradients under coupled transport and reaction. *Geochim. et Cosmochim. Acta* **306**, 189 (2021).
- Tong, Y. *et al.* Exploring the utility of compound-specific isotope analysis for assessing ferrous iron-mediated reduction of RDX in the subsurface. *Environ. Sci. Technol.* **55**, 6752–6763 (2021).
- Li, L., Steefel, C. I. & Yang, L. Scale dependence of mineral dissolution rates within single pores and fractures. *Geochim. et Cosmochim. Acta* **72**, 360 (2008).
- Jung, H. & Meile, C. Upscaling of microbially driven first-order reactions in heterogeneous porous media. *J. Contam. Hydrol.* **224**, 103483 (2019).
- Wen, H. & Li, L. An upscaled rate law for mineral dissolution in heterogeneous media: The role of time and length scales. *Geochim. et Cosmochim. Acta* **235**, 1–20 (2018).
- 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. *Adv. Water Resour.* **123**, 1–11 (2019).
- Atchley, A. L., Navarre-Sitchler, A. K. & Maxwell, R. M. The effects of physical and geochemical heterogeneities on hydrogeochemical transport and effective reaction rates. *J. Contam. Hydrol.* **165**, 53 (2014).
- Min, T., Gao, Y., Chen, L., Kang, Q. & Tao, W. Changes in porosity, permeability and surface area during rock dissolution: Effects of mineralogical heterogeneity. *Int. J. Heat Mass Transf.* **103**, 900 (2016).
- 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 Resour. Res.* **53**, 3645 (2017).
- Liu, M., Shabaninejad, M. & Mostaghimi, P. Impact of mineralogical heterogeneity on reactive transport modelling. *Comput. Geosci.* **104**, 12 (2017).
- Liu, M., Shabaninejad, M. & Mostaghimi, P. Predictions of permeability, surface area and average dissolution rate during reactive transport in multi-mineral rocks. *J. Pet. Sci. Eng.* **170**, 130 (2018).
- 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 Resour. Res.* **55**, 4186–4206 (2019).
- Spokas, K., Peters, C. A. & Pyrak-Nolte, L. Influence of rock mineralogy on reactive fracture evolution in carbonate-rich caprocks. *Environ. Sci. Technol.* **52**, 10144–10152 (2018).
- 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 Resour. Res.* <https://doi.org/10.1029/2011WR011404> (2012).
- 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. *Geochim. et Cosmochim. Acta* **239**, 374 (2018).
- Mostaghimi, P., Liu, M. & Arns, C. H. Numerical simulation of reactive transport on micro-CT images. *Math. Geosci.* **48**, 963 (2016).
- Liu, M. & Mostaghimi, P. Characterisation of reactive transport in pore-scale correlated porous media. *Chem. Eng. Sci.* **173**, 121 (2017).
- Beckingham, L. E. *et al.* Evaluation of accessible mineral surface areas for improved prediction of mineral reaction rates in porous media. *Geochim. et Cosmochim. Acta* **205**, 31 (2017).
- Heyman, J., Lester, D. R., Turuban, R., Méheust, Y. & le Borgne, T. Stretching and folding sustain microscale chemical gradients in porous media. *Proc. Natl. Acad. Sci.* **117**, 13359 (2020).
- Yang, Y. *et al.* Dynamic pore-scale dissolution by CO<sub>2</sub>-saturated brine in carbonates: Impact of homogeneous versus fractured versus vuggy pore structure. *Water Resour. Res.* <https://doi.org/10.1029/2019WR026112> (2020).
- Yoon, S. & Kang, P. K. Roughness, inertia, and diffusion effects on anomalous transport in rough channel flows. *Phys. Rev. Fluids* **6**, 014502 (2021).
- Kanavas, Z., Pérez-Reche, F. J., Arns, F. & Morales, V. L. Flow path resistance in heterogeneous porous media recast into a graph-theory problem. *Transp. Porous Media.* <https://doi.org/10.1007/s11242-021-01671-6> (2021).

41. de Anna, P., Quaife, B., Biros, G. & Juanes, R. Prediction of the low-velocity distribution from the pore structure in simple porous media. *Phys. Rev. Fluids* **2**, 124103 (2017).
42. Boon, M., Bijeljic, B. & Krevor, S. Observations of the impact of rock heterogeneity on solute spreading and mixing. *Water Resour. Res.* **53**, 4624 (2017).
43. Dentz, M., le Borgne, T., Englert, A. & Bijeljic, B. Mixing, spreading and reaction in heterogeneous media: A brief review. *J. Contam. Hydrol.* **120–121**, 1–17 (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. *Transp. Porous Media* **115**, 215 (2016).
45. Liu, M. & Mostaghimi, P. Pore-scale simulation of dissolution-induced variations in rock mechanical properties. *Int. J. Heat Mass Transf.* **111**, 842 (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. *J. Hydrol.* **517**, 1049 (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 Resour. Res.* **50**, 9343 (2014).
48. Jiménez-Martínez, J. *et al.* Homogenization of dissolution and enhanced precipitation induced by bubbles in multiphase flow systems. *Geophys. Res. Lett.* <https://doi.org/10.1029/2020GL087163> (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. *Comput. Geosci.* **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 Resour. Res.* **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. *Environ. Sci. Technol.* **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. *Adv. Water Resour.* **51**, 217 (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. *Int. J. Hydrogen Energy* **40**, 15663–15671 (2015).
54. Noiriel, C. & Soulaine, C. Pore-scale imaging and modelling of reactive flow in evolving porous media: Tracking the dynamics of the fluid–rock interface. *Transp. 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. *Transp. Porous Media* **131**, 427 (2020).
56. Chung, T., da Wang, Y., Armstrong, R. T. & Mostaghimi, P. Minimising the impact of sub-resolution features on fluid flow simulation in porous media. *J. Pet. Sci. Eng.* **207**, 109055 (2021).
57. da Wang, Y., Chung, T., Armstrong, R. T., McClure, J. E. & Mostaghimi, P. Computations of permeability of large rock images by dual grid domain decomposition. *Adv. Water Resour.* **126**, 1–14 (2019).
58. Russell, S. & Norvig, P. *Artificial Intelligence: A Modern Approach* (Prentice Hall, 2020).
59. da Wang, Y., Blunt, M. J., Armstrong, R. T. & Mostaghimi, P. Deep learning in pore scale imaging and modeling. *Earth-Sci. Rev.* **215**, 103555 (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 Resour. Res.* <https://doi.org/10.1029/2020WR029472> (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. *Sci. Rep.* <https://doi.org/10.1038/s41598-021-82029-2> (2021).
62. Jordan, M. I. & Mitchell, T. M. Machine learning: Trends, perspectives, and prospects. *Science* **349**, 255 (2015).
63. Alqhtani, N., Alzubaidi, F., Armstrong, R. T., Swietojanski, P. & Mostaghimi, P. Machine learning for predicting properties of porous media from 2d X-ray images. *J. Pet. Sci. Eng.* **184**, 106514 (2020).
64. Rabbani, A., Babaei, M., Shams, R., da Wang, Y. & Chung, T. DeepPore: A deep learning workflow for rapid and comprehensive characterization of porous materials. *Adv. Water Resour.* **146**, 103787 (2020).
65. Chung, T., da Wang, Y., Armstrong, R. T. & Mostaghimi, P. Voxel agglomeration for accelerated estimation of permeability from micro-CT images. *J. Pet. Sci. Eng.* **184**, 106577 (2020).
66. Leal, A. M. M., Kyas, S., Kulik, D. A. & Saar, M. O. Accelerating reactive transport modeling: On-demand machine learning algorithm for chemical equilibrium calculations. *Transp. Porous Media* **133**, 161 (2020).
67. Guérillot, D. & Bruyelle, J. Geochemical equilibrium determination using an artificial neural network in compositional reservoir flow simulation. *Comput. Geosci.* **24**, 697 (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). <https://doi.org/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. *Sci. Rep.* <https://doi.org/10.1038/s41598-018-34833-6> (2018).
70. Liu, M. & Mostaghimi, P. High-resolution pore-scale simulation of dissolution in porous media. *Chem. Eng. Sci.* **161**, 360 (2017).
71. Liu, M. & Mostaghimi, P. Numerical simulation of fluid–fluid–solid reactions in porous media. *Int. J. Heat Mass Transf.* **120**, 194 (2018).
72. Liu, M. & Mostaghimi, P. Reactive transport modelling in dual porosity media. *Chem. Eng. Sci.* **190**, 436 (2018).
73. Sahimi, M. *Flow and Transport in Porous Media and Fractured Rock: From Classical Methods to Modern Approaches* (Wiley, 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. *Chem. Eng. Sci.* **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 Resour. Res.* **52**, 7198–7212 (2016).
77. Ryan, E. M., Tartakovsky, A. M. & Amon, C. Pore-scale modeling of competitive adsorption in porous media. *J. Contam. Hydrol.* **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. *Environ. Sci. Pollut. Res.* **26**, 17787 (2019).
79. Lasaga, A. C. Chemical kinetics of water–rock interactions. *J. Geophys. Res. Solid Earth* **89**, 4009–4025 (1984).
80. Kang, Q., Zhang, D. & Chen, S. Simulation of dissolution and precipitation in porous media. *J. Geophys. Res. Solid Earth.* <https://doi.org/10.1029/2003JB002504> (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. *Int. Commun. Heat Mass Transf.* **116**, 104694 (2020).
83. Breiman, L. Random Forests. *Mach. Learn.* **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. *J. Guid. Control Dyn.* **13**, 926 (1990).
87. Bebis, G. & Georgiopoulos, M. Feed-forward neural networks. *IEEE Potentials* **13**, 27 (1994).
88. Lei, X., Liao, X., Chen, F. & Huang, T. Two-layer tree-connected feed-forward neural network model for neural cryptography. *Phys. Rev. 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. *Chem. Phys. Lett.* **629**, 40 (2015).
90. Mohri, M., Rostamizadeh, A. & Talwalkar, A. *Foundations of Machine Learning* (MIT Press, 2018).
91. di Buccianico, A. Coefficient of determination ( $R^2$ ). In *Encyclopedia of Statistics in Quality and Reliability* (eds Ruggeri, F. et al.) (Wiley, 2008).
92. Rodriguez, J. D., Perez, A. & Lozano, J. A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 569 (2010).
93. Brown, C. E. Coefficient of variation. In *Applied Multivariate Statistics in Geohydrology and Related Sciences* (ed. Brown, C. E.) (Springer, 1998).
94. Cunningham, P. Dimension reduction. In *Machine Learning Techniques for Multimedia* (eds Cord, M. & Cunningham, P.) (Springer, 2008).
95. Panthong, R. & Srivihok, A. Wrapper feature subset selection for dimension reduction based on ensemble learning algorithm. *Procedia Comput. Sci.* **72**, 162 (2015).
96. Chizi, B. & Maimon, O. Dimension reduction and feature selection. In *Data Mining and Knowledge Discovery Handbook* (eds Maimon, O. & Rokach, L.) (Springer, 2009).
97. Schmitt, M., Halisch, M., Müller, C. & Fernandes, 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. *J. Mater. Sci.* **53**, 5063 (2018).
99. Blunt, M. J. et al. Pore-scale imaging and modelling. *Adv. Water Resour.* **51**, 197 (2013).
100. Gouze, P. & Luquot, L. X-ray microtomography characterization of porosity, permeability and reactive surface changes during dissolution. *J. Contam. Hydrol.* **120–121**, 45–55 (2011).
101. Noiriél, C. et al. Changes in reactive surface area during limestone dissolution: An experimental and modelling study. *Chem. Geol.* **265**, 160–170 (2009).
102. Xie, X. et al. Testing and validating machine learning classifiers by metamorphic testing. *J. Syst. Softw.* **84**, 544 (2011).
103. Esteves, B. F., Lage, P. L. C., Couto, P. & Kovscek, A. R. Pore-network modeling of single-phase reactive transport and dissolution pattern evaluation. *Adv. Water Resour.* **145**, 103741 (2020).
104. You, J. & Lee, K. J. Pore-scale study to analyze the impacts of porous media heterogeneity on mineral dissolution and acid transport using Darcy–Brinkmann–Stokes Method. *Transp. Porous Media* **137**, 575–602 (2021).
105. Salamat, Y. & Hidrovo, C. H. A parametric study of multiscale transport phenomena and performance characteristics of capacitive deionization systems. *Desalination* **438**, 24–36 (2018).
106. Meakin, P. & Tartakovsky, A. M. Modeling and simulation of pore-scale multiphase fluid flow and reactive transport in fractured and porous media. *Rev. Geophys.* **47**, 3002 (2009).

## 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.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-09495-0>.

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

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022