Unveiling the Re, Cr, and I diffusion in saturated compacted bentonite using machine-learning methods

Authors: Zheng-Ye Feng, Jun-Lei Tian, Tao Wu, Guo-Jun Wei, Zhi-Long Li, Xiao-Qiong Shi, Yong-Jia Wang, Qing-Feng Li, Tao Wu

Date: 2024-04-08T00:00:00+00:00

Abstract

The safety assessment of high-level radioactive waste repositories requires a high predictive accuracy for radionuclide diffusion and a comprehensive understanding of the diffusion mechanism. In this study, a through-diffusion method and six machine-learning methods were employed to investigate the diffusion of ReO4-, HCrO4-, and I-in saturated compacted bentonite under different salinities and compacted dry densities. The machine-learning models were trained using two datasets. One dataset contained six input features and 293 instances obtained from the diffusion database system of the Japan Atomic Energy Agency (JAEA-DDB) and 15 publications. The other dataset, comprising 15,000 pseudo-instances, was produced using a multi-porosity model and contained eight input features. The results indicate that the former dataset yielded a higher predictive accuracy than the latter. Light gradient-boosting exhibited a higher prediction accuracy (R2= 0.92) and lower error (MSE = 0.01) than the other machine-learning algorithms. In addition, Shapley Additive Explanations,Feature Importance, and Partial Dependence Plot analysis results indicate that the rock capacity factor and compacted dry density had the two most significant effects on predicting the effective diffusion coefficient, thereby offering valuable insights.

Full Text

Unveiling Re, Cr, and I Diffusion in Saturated Compacted Bentonite Using Machine-Learning Methods

Zheng-Ye Feng[1], Jun-Lei Tian[1], Tao Wu[1],†, Guo-Jun Wei[1],[2], Zhi-Long Li[1],[3], Xiao-Qiong Shi[1], Yong-Jia Wang[1], Qing-Feng Li[1]

[1]Department of Engineering, Huzhou University, Huzhou 313000, China
[2]Institute of Theoretical Physics, Shanxi University, Taiyuan 030006, China
[3]College of Physics Science and Technology, Shenyang Normal University, Shenyang 110034, China

Abstract

The safety assessment of high-level radioactive waste repositories demands high predictive accuracy for radionuclide diffusion and a comprehensive understanding of the underlying diffusion mechanisms. This study employed a through-diffusion method and six machine-learning approaches to investigate the diffusion of ReO_4^-, $HCrO_4^-$, and I^- in saturated compacted bentonite under varying salinities and compacted dry densities. Machine-learning models were trained using two distinct datasets. The first contained six input features and 293 instances compiled from the Japan Atomic Energy Agency diffusion database system (JAEA-DDB) and 15 publications. The second dataset comprised 15,000 pseudo-instances generated using a multi-porosity model with eight input features. Results indicate that the former dataset yielded superior predictive accuracy. Light gradient-boosting demonstrated higher prediction accuracy ($R^2 = 0.92$) and lower error (MSE = 0.01) compared to other machine-learning algorithms. Furthermore, Shapley Additive Explanations, Feature Importance, and Partial Dependence Plot analyses revealed that the rock capacity factor and compacted dry density exerted the two most significant effects on predicting the effective diffusion coefficient, offering valuable mechanistic insights.

Keywords: Machine learning; Effective diffusion coefficient; Through-diffusion experiment; Multi-porosity model; Global analysis

Introduction

China is planning to construct a deep geological repository for high-level radioactive waste in the Beishan area of Gansu Province [?]. Gaomiaozi (GMZ) bentonite from Inner Mongolia has been selected as an engineering barrier material due to its high adsorption capacity, low permeability, favorable thermal conductivity, and abundant reserves [?]. This porous clay mineral possesses a layered structure consisting of tetrahedral-octahedral-tetrahedral sheets. Diffusion represents the primary transport mechanism for radionuclides through the bentonite barrier [?]. Anionic radionuclides with long half-lives, such as $^{129}I^-$, $^{36}Cl^-$, $^{79}SeO_3^{2-}$, and $^{99}TcO_4^-$, are widely recognized as significant contributors to potential long-term dose due to their high diffusivity, which arises from the anionic exclusion effect at the negatively charged bentonite surface [?, ?]. Consequently, evaluating the release of anionic radionuclides from bentonite barriers is crucial for repository safety assessment.

Among diffusion parameters, the effective diffusion coefficient constitutes a critical parameter in safety assessments. It is influenced by numerous factors, including porosity, the species diffusion coefficient in water, radionuclide concentration gradient, and tortuosity [?]. Extensive experimental studies have investigated specific influencing factors such as compacted dry density, ionic strength, bentonite type, and temperature [?]. The relationships between these factors and radionuclide diffusion have been established empirically. For instance, the effective diffusion coefficient increases with decreasing compacted dry density [?] and increasing ionic strength [?, ?]. Bentonites with high montmorillonite content exhibit superior radionuclide retardation due to their lower effective diffusion coefficients [?, ?, ?, ?]. Furthermore, the temperature dependence of the effective diffusion coefficient has been described using the Arrhenius equation [?, ?].

Several numerical models, including the multi-porosity model [?, ?], integrated sorption and diffusion models [?], and pore-scale models [?, ?], have been employed to predict effective diffusion coefficients and analyze the impact of influencing factors. These models have generated theoretical results that align with experimental observations. However, few studies have reported quantitative metrics such as the coefficient of determination (R^2) or mean square error (MSE) to assess predictive accuracy comprehensively.

Machine-learning methods excel at regression analysis and interpreting nonlinear relationships in multi-factor scenarios, making them valuable tools for engineering applications [?, ?]. Numerous studies have utilized machine-learning approaches, including artificial neural networks (ANNs) and gradient-boosting models, to estimate chloride diffusion coefficients in cement [?]. Predictive accuracy can be enhanced by incorporating physical information into the models [?]. These studies have employed techniques such as Individual Conditional Expectation (ICE), Shapley Additive Explanations (SHAP), and Partial Dependence Plots (PDPs) to analyze the relative importance of influencing factors on chloride diffusion [?]. Regression analyses have predicted chloride diffusion coefficients using input features ranging from 4 to 23 and experimental instances ranging from 72 to 843 [?, ?]. Recently, Light Gradient-Boosting (LightGBM) and ANN algorithms were developed to predict the effective diffusion coefficient of Re(VII) using pseudo-instances generated from a multi-porosity model, with the ANN achieving $R^2 = 0.97$ and LightGBM achieving $R^2 = 0.92$ [?]. However, few studies have employed machine-learning models to explain the correlations between influencing factors and radionuclide effective diffusion coefficients.

This study applies machine-learning models to investigate the diffusion of several simulated radionuclide anions—ReO_4^- as an analogue for $^{99}TcO_4^-$, $HCrO_4^-$ as an analogue for certain redox-sensitive monovalent radionuclide anions, and I^- as an analogue for $^{129}I^-$—in compacted bentonite. The predictive accuracy for the effective diffusion coefficient was evaluated using two training datasets: one compiled from the JAEA-DDB and 15 publications, and another containing pseudo-instances generated using the multi-porosity model. The primary objectives are: (i) to enhance the diffusion database by measuring effective diffusion coefficients for ReO_4^-, $HCrO_4^-$, and I^- in compacted bentonite; (ii) to identify high-performance machine-learning algorithms among six models including LightGBM, Extreme Gradient-Boosting (XGBoost), Categorical Gradient-Boosting (Catboost), ANN, Random Forest (RF), and Support Vector Machine

(SVM); and (iii) to assess whether machine-learning models achieve sufficient understanding of diffusion mechanisms through quantitative analysis of influencing factors. The main novelty lies in developing a high-accuracy machine-learning model for radionuclide diffusion prediction and interpreting correlations between influencing factors and effective diffusion coefficients.

Materials and Methods

Materials

Stock solutions of ReO_4^-, $HCrO_4^-$, and I^- were prepared by dissolving weighed amounts of $KReO_4$, $K_2Cr_2O_7$, and NaI in 200 mL of NaCl solution. The initial concentrations were 1.12×10^{-3} mol/L for ReO_4^-, (0.26–2.14) $\times 10^{-3}$ mol/L for $HCrO_4^-$, and 0.04×10^{-3} mol/L for I^-. Concentrations were measured using an Optima 7000DV inductively coupled plasma optical emission spectrometer (ICP-OES, PerkinElmer, USA). All reagents were of analytical grade.

Diffusion Method

A through-diffusion method, which measures diffusion parameters of ions through a defined thickness of porous material, was employed to investigate anion diffusion (ReO_4^-, $HCrO_4^-$, and I^-) in compacted bentonite. Experiments were conducted using 0.10–0.50 mol/L NaCl solutions, with compacted dry densities ranging from 1300 to 1700 kg/m^3, pH maintained at 5.6 ± 0.1, and temperature at 15 ± 3 °C.

Bentonite powder was compacted into blocks ($\Phi 2.1 \times 1.2$ cm) and sandwiched between two stainless-steel filters ($\Phi 2.54 \times 0.1$ cm). The assembly was inserted into a cylindrical cell. After saturating the bentonite blocks with 0.10–0.50 mol/L NaCl solution for one month, the reservoir connected to one side of the diffusion cell (x = 0) was replaced with 200 mL of the prepared stock solution containing ReO_4^-, $HCrO_4^-$, and I^-. The opposite side (x = L) was connected to a target reservoir filled with 10 mL of NaCl solution, which was periodically replaced to maintain a low anion concentration gradient (<5% of the concentration at x = 0). Detailed equipment and procedural descriptions are available in the literature [?].

Self-programmed Fitting for Diffusion Parameters software calculated the rock capacity factor and effective diffusion coefficient by analyzing accumulated mass as a function of time. Parameter reliability was evaluated by examining consistency between calculated and experimental flux results.

Multi-Porosity Model

GMZ and Anji bentonite powders were obtained from Gaomiaozi, Inner Mongolia, and Anji, Zhejiang Province, respectively. GMZ bentonite has a grain density of 2660 kg/m^3, particle size (d_{50}) of 7.1 µm, cation exchange capacity of 77.3 meq/100 g, and external surface area of 25.6 m^2/g. Its mineral composition is 74.5% montmorillonite, 12 wt% quartz, 7 wt% cristobalite, 4 wt% feldspar, 1 wt% calcite, and 1 wt% kaolinite [?]. Anji bentonite has a particle size (d_{50}) of 11.6 µm, cation exchange capacity of 76 meq/100 g, and external surface area of 60.3 m^2/g, with a mineral composition of 46 wt% montmorillonite, 33 wt% quartz, 10 wt% orthoclase, 8 wt% microcline, and 3 wt% calcite [?].

A multi-porosity model was established for montmorillonite microstructure, as montmorillonite is the predominant mineral in bentonite. The model considers only through-pores in compacted bentonite, where total porosity ( ) is subdivided into three components: diffuse double-layer porosity (_{ddl}), interlayer porosity (_{il}), and free-layer porosity (_{free}) [?, ?]. When compacted bentonite is saturated with aqueous solution, diffuse double-layer pores form transition zones from particle surfaces to free pore water, containing an anion deficit, water molecules, and cation excess. Interlayer pores contain cations and water molecules, with excess cations compensating for charge deficits in the tetrahedral-octahedral-tetrahedral layers, while water molecules arrange in layers [?]. Free-layer pores comprise spaces containing charge-balanced anions, cations, and water molecules.

Due to the anionic exclusion effect, anionic radionuclides cannot readily enter bentonite interlayer pores. Therefore, the model assumes free-layer pores as the predominant diffusion pathways, with accessible porosity defined as:

_{acc} _{free} = _{tot} - _{ddl} - _{il}

The diffuse double-layer porosity _{ddl}, which depends on ionic strength, external surface area, montmorillonite mass ratio, and compacted dry density, is estimated as:

_{ddl} = 3.09 × 10^{-10} A_{ext} · m · _d

Interlayer porosity depends on compacted dry density, water layer fraction, and montmorillonite mass ratio. Interlayer water relates to compaction degree: one water layer ranges between 12.2-12.7 Å at 1600-2000 kg/m^3, two water layers between 15.2-15.7 Å at 1300-1600 kg/m^3, and three water layers between 18.4-19 Å below 1300 kg/m^3 [?]. The interlayer porosity _{il} is approximated as [?]:

- At _d ≤ 1300 kg/m^3: _{il} = m · 1300 · w_3
- At _d > 1300 kg/m^3: _{il} = m · _d · ($x_1 w_1 + x_2 w_2 + x_3 w_3$)

where interlayer water amounts w_i are: 0.119 kg H_2O/kg clay for one water layer at 1600 ≤ _d ≤ 2000 kg/m^3; 0.238 kg H_2O/kg clay for two water layers at 1300 ≤ _d ≤ 1600 kg/m^3; and 0.357 kg H_2O/kg clay for three water layers at _d ≤ 1300 kg/m^3.

The layer fraction x_i is approximately calculated as follows, where subscript i denotes one, two, or three water layers [?, ?]:

- At 1300 ≤ _d ≤ 1600 kg/m^3: $x_1 = 0$, x_2 = (_d - _d,3WL$$\rightarrow 2WL)/(\rho_d, 2WL\rightarrow 1WL - \rho_d, 3WL\rightarrow 2WL), x\_\{3\}$$ = 1 - x_2

- At 1600 ≤ _d ≤ 2000 kg/m^3: x_1 = (_d - _d,2WL$\rightarrow 1WL)/(\rho_d, 1WL - \rho_d, 2WL\rightarrow 1WL), x_{2}$ = 1 - x_1, x_3 = 0

The effective diffusion coefficient is estimated by combining the accessible porosity with Archie' s equation:

D_e = D_w · _{acc}^{n} = D_w · (_{tot} - _{ddl} - _{il})^n

where D_w denotes the species diffusion coefficient in water (1.46×10^{-9} m^2/s for ReO_4^-, 1.13×10^{-9} m^2/s for $HCrO_4^-$, and 2.0×10^{-9} m^2/s for I^- [?]), and n(-) represents the cementation factor.

Database Description and Analysis

The training dataset was obtained from two sources. The first comprised experimental instances from JAEA-DDB (223 instances, 1989-2005) [?] and 15 publications (99 instances, 2006-2024), listed in Table S1 (Supporting Information). The second database contained 15,000 pseudo-instances generated using the multi-porosity model [?]. Table 1 summarizes the statistical information for both datasets. For the JAEA-DDB and literature dataset, only anion diffusion instances in bentonite were selected.

Data preprocessing employed Mahalanobis distance (MD) to remove outliers. MD is a multivariate distance measure accounting for data mean and covariance. The cutoff point d_i is defined as:

d_i = √[(X_i - X̄) · C^{-1} · (X_i - X̄)]

where C, X_i, and X̄ represent the sample covariance matrix, object vector, and arithmetic mean vector, respectively. A cutoff point of five identified 29 outliers, yielding final datasets of 197 JAEA-DDB instances and 96 publication instances for machine-learning models.

Input features for the JAEA-DDB/publications dataset were: rock capacity factor, compacted dry density, montmorillonite mass ratio, species diffusion coefficient in water, ionic strength, and temperature. The multi-porosity model dataset included: external surface area, montmorillonite mass ratio, ionic strength, accessible porosity, compacted dry density, cementation factor, fitting parameter, and species diffusion coefficient in water. The rock capacity factor indicates bentonite' s ability to impede radionuclide diffusion into granite rock, equaling accessible porosity when below total porosity. External surface area, accessible porosity, and cementation factor characterize bentonite properties, while the species diffusion coefficient represents radionuclide properties. Temperature, ionic strength, and compacted dry density are experimental condition parameters. The effective diffusion coefficient served as the sole output feature.

The test dataset comprised eight instances from through-diffusion experiments measuring ReO_4^-, $HCrO_4^-$, and I^- diffusion. Since both effective and species diffusion coefficients ranged from 10^{-13} to 10^{-9} m^2/s, logarithmic conversion was applied to maintain consistency with other features spanning 0–2000, improving model performance [?].

Performance Evaluation of Machine-Learning Models

Predictive accuracy was evaluated using R^2 and MSE, calculated as:

R^2 = 1 - [Σ(log D_{exp}e,i - log D_{pred}e,i)^2] / [Σ(log D_{exp}e,i - log D_{exp}e,ave)^2]

MSE = (1/N) Σ(log D_{exp}e,i - log D_{pred}e,i)^2

where N is the number of instances, log D_{exp}e,i and log D_{pred}e,i represent experimental and predicted values, respectively, and log D_{exp}e,ave denotes the experimental average. Higher R^2 and lower MSE indicate better predictive accuracy.

Five-fold cross-validation (CV) mitigated overfitting, where the dataset was randomly divided into five equal subsamples, with four used for training and one for testing in each iteration.

Results and Discussion

Database Distribution and Characteristics

Figures 1A–1F show the dependence of the effective diffusion coefficient on each input feature for the JAEA-DDB/publications dataset (multi-porosity model dependencies are reported elsewhere [?]). Histograms and kernel density curves display input feature and effective diffusion coefficient distributions. Curve shapes reflect data point concentration; higher concentrations produce greater peak amplitudes.

The rock capacity factor can be obtained directly from through-diffusion experiments or calculated as:

α = + _d · K_d

where (-), _d (kg/m^3), and K_d (m^3/kg) represent porosity, compacted dry density, and distribution coefficient, respectively. Total porosity of compacted bentonite equals that of neutral molecules like HTO [?, ?, ?], while accessible porosity approximates that of anionic radionuclides such as $^{36}Cl^-$ and $^{125}I^-$ [?], based on the ionic exclusion effect hindering anion access to negatively charged surfaces [?, ?]. The rock capacity factor ranged from 0.01 to 19.08 (Table 1), with most data concentrated below two: 13.3% exceeded two, 23.2% ranged from one to two, and 63.5% were less than one. Only nine data points exceeded ten (Fig. 1A). Oscarson et al. [?] reported $^{99}TcO_4^-$ and $^{125}I^-$ rock capacity factors >5 in 4.9% of cases, possibly due to overestimated distribution coefficients from sorption experiments with powdered bentonite.

The effective diffusion coefficient increased with decreasing compacted dry density (Fig. 1B), consistent with previous experimental results [?]. It also increased with species diffusion coefficient in water and temperature (Figs. 1C and 1D), following Archie' s law (D_e = D_w · ^{n}) [?, ?] and the Arrhenius equation (D_e = A ·e^(-E_a/RT)) [?]. Figures 1E and 1F show smeared distributions for concentrated data within limited ranges, indicating no strict one-to-one dependency on montmorillonite mass ratio or ionic strength.

Measurement of Diffusion Parameters Using the Through-Diffusion Method

The through-diffusion method determined diffusion parameters for ReO_4^-, $HCrO_4^-$, and I^- in compacted bentonite. Figure 2 [Figure 2: see original paper] presents breakthrough curves under various salinity and compaction conditions. Salinity effects in GMZ bentonite are shown in Figs. 2A–2E, while dry density effects in Anji bentonite appear in Figs. 2F–2H. Red dots and lines represent flux results; blue dots and lines show accumulated mass. Solid dots indicate experimental data; lines represent calculated temporal relationships for accumulated mass or flux; shaded areas denote upper and lower uncertainty limits considering rock capacity factor and effective diffusion coefficient uncertainties from sample weight, bentonite block and filter volumes, diffusion cell dead volume, and ICP-OES measurements. Ionic strength I (mol/L) was calculated as I = 0.5 Σ C_i z_i^2, where C_i is species concentration and z_i is charge number.

Table 2 lists diffusion parameters for ReO_4^-, $HCrO_4^-$, and I^- in compacted bentonite. Both rock capacity factor and effective diffusion coefficient increase with ionic strength and decrease with compacted dry density, consistent with previous studies [?, ?, ?]. ReO_4^- exhibited higher effective diffusion coefficients than previous GMZ bentonite studies [?], attributable to lower compacted dry densities investigated here. Higher effective diffusion coefficients for $HCrO_4^-$ and I^- compared to literature values [?, ?] resulted from higher ionic strength, lower montmorillonite mass ratio in $HCrO_4^-$ experiments, and higher ionic strength and dry density in I^- experiments. These results comply with anion diffusion laws and fall within the training dataset range, validating the test dataset for evaluation. Notably, measured rock capacity factors for ReO_4^-, $HCrO_4^-$, and I^- are less than total porosity, indicating no adsorption and equivalence to accessible porosity.

Prediction by Machine-Learning Algorithms

Six machine-learning algorithms—LightGBM, XGBoost, Catboost, ANN, RF, and SVM—predicted effective diffusion coefficients using two training datasets. The first contained eight input features and 15,000 pseudo-instances from the multi-porosity model. The second included six input features and 293 instances from JAEA-DDB and publications (Table 1). Datasets were split 4:1 for training and validation. The test dataset comprised experimental results from Table 2.

LightGBM, XGBoost, Catboost, and RF are ensemble-learning algorithms; ANN and SVM are traditional learning algorithms. Table 3 lists mean performance metrics from five-fold CV. LightGBM achieved the highest predictive performance with R^2{CV} = 0.87 and MSE{CV} = 0.01.

Hyperparameters, which control model learning and cannot be learned from data, were tuned using grid search (GS). Reasonable hyperparameter ranges were manually predefined, and the model iterated through all combinations. Cross-validation guided selection for training datasets, yielding optimal parameter combinations. Table 4 summarizes tuned hyperparameters.

Figure 3 [Figure 3: see original paper] compares experimental and predicted effective diffusion coefficients. Dots represent experimental data; red lines show linear fits; shaded areas indicate 95% confidence intervals. For the multi-porosity model dataset, predictive accuracy ranked: SVM > XGBoost > RF > LightGBM > CatBoost > ANN (Figs. 3A–3F). SVM achieved MSE = 0.01 and $R^2 = 0.83$. For the JAEA-DDB/publications dataset, accuracy ranked: LightGBM > CatBoost > XGBoost > SVM > RF > ANN (Figs. 3G–3L). All gradient-boosting algorithms achieved $R^2 > 0.88$, with LightGBM and XGBoost reaching MSE = 0.01 and $R^2 = 0.91$. The JAEA-DDB/publications dataset outperformed the multi-porosity model dataset, likely due to the complexity of predicting multiple species (ReO_4^-, $HCrO_4^-$, I^-) under varying salinity and compaction conditions, which challenges models trained on multi-porosity model data. Boosting models combining weak learners via weight-based aggregation showed stronger predictive capabilities, consistent with previous studies [?, ?]. LightGBM achieved highest accuracy among boosting models through gradient-based one-side sampling and exclusive feature bundling techniques [?, ?].

Shapley Additive Explanation and Feature Importance Analyses

SHAP and Feature Importance (FI) are widely used feature attribution methods that identify input feature significance driving predictions [?]. Though employing distinct techniques, both rank feature importance by influence on predicted output [?]. Applied to the highest-performing LightGBM model with the JAEA-DDB/publications dataset, both analyses identified rock capacity factor and compacted dry density as the two most important features (Fig. 4 [Figure 4: see original paper]). For remaining features, FI ranked: T > log D_w > I > m, while SHAP ranked: log D_w T > m > I. The montmorillonite mass ratio ranking difference stems from underlying methodological principles.

Ionic strength is closely associated with the electrical double layer at bentonite interfaces [?]. Despite limited effect on effective diffusion coefficient prediction (Fig. 4), its influence has been extensively investigated experimentally [?, ?, ?]. The effective diffusion coefficient increases with salinity until ionic strength exceeds 0.5 mol/L, explained by minimum electrical double-layer thickness, negligible diffuse double-layer pores, and maximum free-layer pore width [?, ?, ?]. Debate continues regarding the electrical double-layer effect on radionuclide diffusion [?, ?], potentially due to the small porosity proportion in diffuse double-layer pores [?]. Feature importance depends on input features, instances, and algorithms, warranting further research to clarify ionic strength significance.

Partial Dependence Plot Analysis

Partial Dependence Plot (PDP) analysis quantifies relationships between input and output features, assessing positive and negative effects on the effective diffusion coefficient (Fig. 5 [Figure 5: see original paper]). Features strongly impacting the output exhibit significant PDP curve changes, while low-impact features produce flat curves.

Rock capacity factor, species diffusion coefficient in water, ionic strength, and temperature positively impacted the effective diffusion coefficient, whereas compacted dry density and montmorillonite mass ratio negatively impacted it. The effective diffusion coefficient increases with species diffusion coefficient, temperature, and ionic strength, consistent with Archie' s law [?, ?], the Arrhenius equation [?], and experimental results [?, ?]. Conversely, it decreases with high compacted dry density and montmorillonite mass ratio, also consistent with previous studies [?, ?].

The rock capacity factor exhibited the strongest influence. Increasing from 0.01 to 19.08 raised the PDP value from -11.16 to -9.90, representing an ~11.3% increase (Fig. 5A). For anionic radionuclides, rock capacity factors should remain below total porosity when considering anionic exclusion [?, ?], suggesting some anomalous instances in JAEA-DDB should theoretically be removed, though they were retained for database integrity. PDP value increases ranked: α (11.3%) > T (8.8%) > log D_w (6.5%) > I (1.3%).

Compacted dry density showed negative impact: increasing from 400 to 1700 kg/m^3 decreased the PDP value from -9.85 to -10.74 (~9.0% decrease) (Fig. 5B), consistent with literature [?]. Montmorillonite mass ratio also negatively impacted: increasing from 0.33 to 1.0 decreased the PDP value from -10.02 to -10.31 (~2.8% decrease) (Fig. 5E). Bentonites with low montmorillonite content, such as illite/smectite mixed-layer (m = 0.33) and Kunigel V1 (m = 0.46-0.49), exhibit higher effective diffusion coefficients [?, ?, ?, ?, ?], confirming that higher montmorillonite content improves radionuclide blocking capacity [?]. Figures 5D and 5F show predicted effective diffusion coefficients increase with ionic strength and temperature, with significant effects observed from 0.01-0.6 mol/L and 22-60 °C, respectively, consistent with previous findings [?, ?]. Thus, PDP analysis provides interpretability of diffusion laws and mechanisms.

Conclusion

Effective diffusion coefficients of ReO_4^-, $HCrO_4^-$, and I^- in compacted Gaomiaozi and Anji bentonites were investigated under various ionic strength and dry density conditions using through-diffusion experiments and six machine-learning models. Key findings are:

(i) Training and validation datasets from experimental instances outperformed pseudo-instance datasets.

(ii) Light Gradient-Boosting achieved higher predictive accuracy than other algorithms, with MSE = 0.01 and R^2 = 0.92 for the JAEA-DDB/publications dataset.

(iii) SHAP, Feature Importance, and PDP analyses identified rock capacity factor and compacted dry density as the two most important features, with positive and negative influences, respectively.

This study introduces a novel high-accuracy machine-learning model for radionuclide diffusion prediction and explores diffusion mechanisms by ranking influencing factors and analyzing their dependencies. Machine-learning algorithms offer powerful tools and a new paradigm for studying radioactive anion diffusion in bentonite barriers. Future research should evaluate model improvement by incorporating additional bentonite characteristic parameters, complex chemical species, and broader geochemical conditions relevant to high-level radioactive waste repositories.

Data Availability

The data used in this study are available online (https://10.5281/zenodo.10837576).

References

[1] Z. Y. Chen, S. Y. Wang, H. J. Hou et al., China' s progress in radionuclide migration study over the past decade (2010–2021): Sorption, transport and radioactive colloid. Chin. Chem. Lett. 33, 3405–3412 (2022). https://doi.org/10.1016/j.cclet.2022.02.054

[2] H. Liu, T. Fu, M. T. Sarwar et al., Recent progress in radionuclides adsorption by bentonite-based materials as ideal adsorbents and buffer/backfill materials. Appl. Clay Sci. 232, 106796 (2023). https://doi.org/10.1016/j.clay.2022.106796

[3] Z. Sun, Y. Chen, Y. Cui et al., Effect of synthetic water and cement solutions on the swelling pressure of compacted Gaomiaozi(GMZ) bentonite: The Beishan site case, Gansu, China. Eng. Geol. 244, 66–74 (2018). https://doi.org/10.1016/j.enggeo.2018.08.002

[4] L. Y. Cui, S. A. Masum, W. M. Ye et al., Investigation on gas migration behaviours in saturated compacted bentonite under rigid boundary conditions. Acta Geotech. 17, 2517–2531 (2022). https://doi.org/10.1007/s11440-021-01424-1

[5] C. D. Shackelford, S. M. Moore, Fickian diffusion of radionuclides for engineered containment barriers: Diffusion coefficients, porosities, and complicating issues. Eng. Geol. 152, 133–147 (2013). https://doi.org/10.1016/j.enggeo.2012.10.014

[6] R. V. H. Dagnelie, P. Arnoux, J. Radwan et al., Perturbation induced by EDTA on HDO, Br^- and EuIII diffusion in a large-scale clay rock sample. Appl. Clay Sci. 105–106, 142–149 (2015). https://doi.org/10.1016/j.clay.2014.12.004

[7] C. Tournassat, C. A. J. Appelo, Modelling approaches for anion-exclusion in compacted Na-bentonite. Geochim. Cosmochim. Acta. 75, 3698–3710 (2011). https://doi.org/10.1016/j.gca.2011.04.001

[8] E. Tosoni, A. Salo, J. Govaerts et al., Comprehensiveness of scenarios in the safety assessment of nuclear waste repositories. Reliab. Eng. Syst. Saf. 188, 561–573 (2019). https://doi.org/10.1016/j.ress.2019.04.012

[9] T. Wu, Y. Yang, Z. Wang et al., Anion diffusion in compacted clays by pore-scale simulation and experiments. Water Resour. Res. 56, 2019WR027037 (2020). https://doi.org/10.1029/2019wr027037

[10] C. Wittebroodt, S. Savoye, B. Frasca et al., Diffusion of HTO, $^{36}Cl^-$ and $^{125}I^-$ in upper toarcian argillite samples from Tournemire: Effects of iodide concentration and ionic strength. Appl. Geochem. 27, 1432–1441 (2012). https://doi.org/10.1016/j.apgeochem.2011.12.017

[11] Y. Tachi, K. Yotsuji, Diffusion and sorption of Cs^+, Na^+, I^- and HTO in compacted sodium montmorillonite as a function of porewater salinity: Integrated sorption and diffusion model. Geochim. Cosmochim. Acta. 132, 75–93 (2014). https://doi.org/10.1016/j.gca.2014.02.004

[12] L. R. Van Loon, J. M. Soler, M. H. Bradbury, Diffusion of HTO, $^{36}Cl^-$ and $^{125}I^-$ in Opalinus Clay samples from Mont Terri. J. Contam. Hydrol. 61, 73–83 (2003). https://doi.org/10.1016/S0169-7722(02)00114-6

[13] M. Bestel, M. A. Glaus, S. Frick et al., Combined through-diffusion of HTO and tracer through-diffusion of $^{22}Na^+$ in Na-montmorillonite with different dry densities. Appl. Geochem. 98, 1-9 (2018). https://doi.org/10.1016/j.apgeochem.2018.04.008

[14] T. Kozaki, A. Fujishima, N. Saito et al., Effects of dry density and exchangeable cations on the diffusion process of sodium ions in compacted montmorillonite. Eng. Geol. 81, 246–254 (2005). https://doi.org/10.1016/j.enggeo.2005.06.010

[15] M. Molera, T. Eriksen, M. Jansson, Anion diffusion pathways in bentonite clay compacted to different dry densities. Appl. Clay Sci. 23, 69–76 (2003). https://doi.org/10.1016/s0169-1317(03)00088-7

[16] H. Sato, T. Ashida, Y. Kohara et al., Effect of dry density on diffusion of some radionuclides in compacted sodium bentonite. J. Nucl. Sci. Technol. 29, 872–882 (1992). https://doi.org/10.1080/18811248.1992.9731607

[17] T. Wu, W. Dai, G. Xiao et al., Influence of dry density on HTO diffusion in GMZ bentonite. J. Radioanal. Nucl. Chem. 292, 853–857 (2012). https://doi.org/10.1007/s10967-011-1523-y

[18] T. Wu, J. Li, W. Dai et al., Effect of dry density on ^{125}I diffusion in GMZ bentonite. Sci. China Chem. 55, 1760–1764 (2012).

https://doi.org/10.1007/s11426-012-4695-6

[19] Y. Fukatsu, K. Yotsuji, T. Ohkubo et al., Diffusion of tritiated water, $^{137}Cs^{+}$, and $^{125}I^{-}$ in compacted Ca-montmorillonite: Experimental and modeling approaches. Appl. Clay Sci. 211, 106176 (2021). https://doi.org/10.1016/j.clay.2021.106176

[20] F. González Sánchez, L. R. Van Loon, T. Gimmi et al., Self-diffusion of water and its dependence on temperature and ionic strength in highly compacted montmorillonite, illite and kaolinite. Appl. Geochem. 23, 3840–3851 (2008). https://doi.org/10.1016/j.apgeochem.2008.08.008

[21] S. Savoye, C. Beaucaire, B. Grenut et al., Impact of solution ionic strength on strontium diffusion through the Callovo-Oxfordian clayrocks: An experimental and modeling study. Appl. Geochem. 61, 41–52 (2015). https://doi.org/10.1016/j.apgeochem.2015.05.011

[22] J. M. Soler, C. I. Steefel, T. Gimmi et al., Modeling the ionic strength effect on diffusion in clay: The DR-A Experiment at Mont Terri. ACS Earth Space Chem. 3, 442–451 (2019). https://doi.org/10.1021/acsearthspacechem.8b00192

[23] W. Tian, C. Li, X. Liu et al., The effect of ionic strength on the diffusion of ^{75}Se(IV) in Gaomiaozi bentonite. J. Radioanal. Nucl. Chem. 295, 1423–1430 (2013). https://doi.org/10.1007/s10967-012-2284-y

[24] T. Wu, Z. Wang, H. Wang et al., Salt effects on Re(VII) and Se(IV) diffusion in bentonite. Appl. Clay Sci. 141, 104–110 (2017). https://doi.org/10.1016/j.clay.2017.02.021

[25] M. García-Gutiérrez, J. L. Cormenzana, T. Missana et al., Diffusion coefficients and accessible porosity for HTO and ^{36}Cl in compacted FEBEX bentonite. Appl. Clay Sci. 26, 65–73 (2004). https://doi.org/10.1016/j.clay.2003.09.012

[26] T. Kozaki, J. Liu, S. Sato, Diffusion mechanism of sodium ions in compacted montmorillonite under different NaCl concentration. Phys. Chem. Earth. 33, 957–961 (2008). https://doi.org/10.1016/j.pce.2008.05.007

[27] Z. Feng, Z. Gao, Y. Wang et al., Application of machine learning to study the effective diffusion coefficient of Re(VII) in compacted bentonite. Appl. Clay Sci. 243, 107076 (2023). https://doi.org/10.1016/j.clay.2023.107076

[28] Z. Geng, Z. Feng, H. Li et al., Porosity investigation of compacted bentonite using through-diffusion method and multi-porosity model. Appl. Geochem. 146, 105480 (2022). https://doi.org/10.1016/j.apgeochem.2022.105480

[29] Y. Yang, M. Wang, Cation diffusion in compacted clay: A pore-scale view. Environ. Sci. Technol. 53, 1976–1984 (2019). https://doi.org/10.1021/acs.est.8b05755

[30] H. L. Liu, H. B. Ji, J. M. Zhang et al., Novel algorithm for detection and identification of radioactive materials in an urban environment. Nucl. Sci. Tech. 34, 154 (2023). https://doi.org/10.1007/s41365-023-01304-1

[31] B. Mortazavi, B. Javvaji, F. Shojaei et al., Exceptional piezoelectricity, high thermal conductivity and stiffness and promising photocatalysis in two-dimensional $MoSi_2N_4$ family confirmed by first-principles. Nano Energy. 82, 105716 (2021). https://doi.org/10.1016/j.nanoen.2020.105716

[32] L. Jin, T. Dong, T. Fan et al., Prediction of the chloride diffusivity of recycled aggregate concrete using artificial neural network. Mater. Today Commun. 32, 104137 (2022). https://doi.org/10.1016/j.mtcomm.2022.104137

[33] E. Samaniego, C. Anitescu, S. Goswami et al., An energy approach to the solution of partial differential equations in computational mechanics via machine learning: Concepts, implementation and applications. Comput. Methods Appl. Mech. Eng. 362, 112790 (2020). https://doi.org/10.1016/j.cma.2019.112790

[34] V. Q. Tran, Machine learning approach for investigating chloride diffusion coefficient of concrete containing supplementary cementitious materials. Constr. Build. Mater. 328, 127103 (2022). https://doi.org/10.1016/j.conbuildmat.2022.127103

[35] N. D. Hoang, C. T. Chen, K. W. Liao, Prediction of chloride diffusion in cement mortar using multi-gene genetic programming and multivariate adaptive regression splines. Measurement. 112, 141–149 (2017). https://doi.org/10.1016/j.measurement.2017.08.031

[36] O. A. Hodhod, H. I. Ahmed, Developing an artificial neural network model to evaluate chloride diffusivity in high performance concrete. HBRC J. 9, 15-21 (2013). https://doi.org/10.1016/j.hbrcj.2013.04.001

[37] W. Z. Taffese, L. Espinosa-Leal, A machine learning method for predicting the chloride migration coefficient of concrete. Constr. Build. Mater. 348, 128566 (2022). https://doi.org/10.1016/j.conbuildmat.2022.128566

[38] Z. J. Wen, Selection and basic properties of the buffer material for high-level radioactive waste repository in China. Acta Geol. Sin-Engl. 82, 1050-1055 (2008). https://doi.org/10.1111/j.1755-6724.2008.tb00662.x

[39] T. Wu, Z. Wang, Y. Tong et al., Investigation of Re(VII) diffusion in bentonite by through-diffusion and modeling techniques. Appl. Clay Sci. 166, 223-229 (2018). https://doi.org/10.1016/j.clay.2018.08.023

[40] M. Holmboe, S. Wold, M. Jonsson, Porosity investigation of compacted bentonite using XRD profile modeling. J. Contam. Hydrol. 127, 19–32 (2012). https://doi.org/10.1016/j.jconhyd.2011.10.005

[41] P. Vanysek, Ionic conductivity and diffusion at infinite dilution. In: CRC Handbook of Chemistry and Physics. 83, 76–78 (2000).

[42] Y. Tochigi, Y. Tachi, Development of diffusion database of buffer materials and rocks–expansion and application method of foreign buffer materials. JAEA-Data/Code 2009-029. (2010). Japan Atomic Energy Agency.

[43] L. R. Van Loon, J. Mibus, A modified version of Archie' s law to estimate effective diffusion coefficients of radionuclides in argillaceous rocks and

its application in safety analysis studies. Appl. Geochem. 59, 85–94 (2015). https://doi.org/10.1016/j.apgeochem.2015.04.002

[44] D. W. Oscarson, H. B. Hume, J. W. Choi, Diffusion of ^{14}C in compacted bentonite and crushed granite. Radiochim. Acta. 65, 189–194 (1994). https://doi.org/10.1524/ract.1994.65.3.189

[45] T. Wu, Z. Wang, Q. Li et al., Re(VII) diffusion in bentonite: Effect of organic compounds, pH and temperature. Appl. Clay Sci. 127-128, 45–51 (2016). https://dx.doi.org/10.1016/j.clay.2016.03.039

[46] K. Liu, J. Zheng, F. Pacheco-Torgal et al., Innovative modeling framework of chloride resistance of recycled aggregate concrete using ensemble-machine-learning methods. Constr. Build. Mater. 337, 127613 (2022). https://doi.org/10.1016/j.conbuildmat.2022.127613

[47] Z. Gao, Y. Wang, H. Lü et al., Machine learning the nuclear mass. Nucl. Sci. Tech. 32, 109 (2021). https://doi.org/10.1007/s41365-021-00956-1

[48] Y. Wang, Z. Gao, H. Lü et al., Decoding the nuclear symmetry energy event-by-event in heavy-ion collisions with machine learning. Phys. Lett. B. 835, 137508 (2022). https://doi.org/10.1016/j.physletb.2022.137508

[49] A. Idiart, M. Pekala, Models for diffusion in compacted bentonite. SKB TR-16-06 (2016). Swedish Nuclear Fuel and Waste Management Company.

Note: Figure translations are in progress. See original paper for figures.