



# Machine learning-enabled prediction of chemical durability of $A_2B_2O_7$ pyrochlore and fluorite



Bowen Gong<sup>a,\*</sup>, Kun Yang<sup>a</sup>, James A. Lian<sup>b</sup>, Jianwei Wang<sup>c</sup>

<sup>a</sup> Department of Mechanical, Aerospace, Nuclear Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

<sup>b</sup> Niskayuna High School, Niskayuna, NY 12309, USA

<sup>c</sup> Department of Geology and Geophysics, Louisiana State University, LA 70803, USA

## ARTICLE INFO

### Keywords:

Machine learning  
Nuclear waste forms  
Pyrochlore and fluorite  
Leaching  
Chemical durability

## ABSTRACT

Pyrochlore-structure type and its derivative in a general formula  $A_2B_2O_7$  ( $A$  = rare earth elements and actinides;  $B$  = Ti, Sn, Zr, Hf, Pb, Si, etc.) display excellent structural flexibility and rich crystal chemistry as promising nuclear waste form materials capable of immobilizing actinides and fission products. It is essential to understand these materials' chemical durability and element release of radionuclides in order to evaluate their performance in near-field environment. However, it is a formidable grand technological challenge to experimentally perform durability testing across hundreds of thousands of possibilities resulting from their extreme compositional complexities due to cation substitutions at both A and B-sites. In this work, we demonstrate a machine learning approach to determine the key materials parameters and structural characteristics governing the leaching behaviors from a small set of selected compositions as model systems, enabling a science-based prediction of their chemical durability that can be extended to a wide range of chemical compositions. The combination of four key structural characteristics and materials parameters, including ionic radius size difference  $r_{A-B}$ , ionic potential difference  $E_{B-A}$ , electronegativity difference  $\chi_{B-A}$ , and lattice parameter  $a$ , creates features an optimized prediction of the chemical durability. Two machine learning models, linear regression and Kernel ridge regression models, are trained on the randomly-split training dataset derived from the experimentally-determined elemental release rates, and subsequently tested on the testing dataset. The predicted leaching rates from both machine learning models show an excellent agreement with the experimental data, demonstrating the feasibility of rapidly evaluating the material properties of new compositions. These results highlight the immense potential of synergizing informatics through machine learning-based models and well-controlled experiments of selected model systems to accelerate materials design and discovery with optimized compositions and performance of promising materials for effective nuclear waste management.

## 1. Introduction

The safe disposal of the high-level waste (HLW) [1] from nuclear weapon production and used fuel reprocessing and excess Pu from weapon programs is the most daunting challenge [2] to the environmental remediation due to their very complex waste streams, high radioactivity of over several hundreds of millions of curies and potential hazardous impacts to the geochemical environment. Long-lived actinide and transuranic elements (e.g., 24,100 years for Pu-239 and 2.14 million years for Np-237 [3]) are the major dose contributors after hundreds of thousands of years of waste storage. Among the many different materials proposed for the storage and immobilization of plutonium and HLWs,

borosilicate glasses [4–6] and ceramics [7–9] are the most popular waste forms extensively studied. The liquid HLW can be encapsulated into borosilicate glass as the baseline waste form material destined for permanent geological disposition. Excess Pu and highly-enriched uranium from the weapon program and actinides from chemical processing of used fuels can be either reused as valuable resources in mixed oxide fuels (MOX) or incorporated into crystalline ceramic waste forms for permanent disposition in deep-underground repositories [10].

Materials for waste form applications typically require very complex crystal chemistry, structural flexibility, and thus capabilities to accommodate a wide range of waste elements with different valence states [11]. Waste form materials should have excellent radiation tolerance,

\* Corresponding author.

E-mail address: [gongb@rpi.edu](mailto:gongb@rpi.edu) (B. Gong).

thermal stability, and well-maintained structural integrity in projected geological repository environments [12]. Ceramics with a general chemical form  $VIII_A_2VI_B_2IV_O_7$  ( $A$  = lanthanides or actinides;  $B$  = Ti, Sn, Zr, Hf, etc.) with fluorite or pyrochlore structures are promising waste form materials for actinide immobilization. The crystal structure of pyrochlore  $A_2B_2O_7$  is in a space group  $Fd\cdot3\ m$  with each unit cell containing 8 molecules. Lanthanides and actinides are often occupied at the 8-coordinated A-site; while Zr, Ti, and group 4A elements are often at the 6-coordinated B-site [1].  $A_2B_2O_7$  pyrochlore can also transform to the parent fluorite structure upon the chemical disordering of the A- and B-site cations and randomization of the oxygen sublattice.  $A_2B_2O_7$  pyrochlore and fluorite structure-types display rich crystal chemistry and structural flexibility with coupled substitution at both A- and B-sites. The phase and radiation stability of  $A_2B_2O_7$  structure-types can be manipulated by controlling the relative ionic size difference at both A- and B-sites or their ionic radius ratio ( $r_{B4+}/r_{A3+}$ ). In addition to single-component systems, binary and multicomponent solid solutions occupying at either A- or B-sites such as high entropy pyrochlore ceramics have been demonstrated that are capable of uranium incorporation [13]. The extreme compositional complexity results in several hundreds of thousands of possibilities and enables a large materials design window by controlling cation substitutions and structural characteristics in order to discover promising materials with optimized performance for effective nuclear waste management.

Chemical durability is critical to evaluate the performance of waste form materials under a near-field environment of a geological repository, and it can be generally assessed by elemental release rates through accelerated leaching test protocols such as static leaching (e.g., the product consistency test, semi-dynamic leaching following an ASTM C1308 protocol and dynamic leaching testing (e.g., single-pass flow testing) [14–19]. A science-based approach to evaluating materials' chemical durability is essential to gain a fundamental understanding of degradation and corrosion mechanisms and kinetics, radionuclide confinement and release, and how different material parameters/structural characteristics affect their chemical durability, in order to optimize waste form design and predict their performance [20]. Extensive data in the literature have been reported for the chemical durability of pyrochlore and fluorite structural types in a general formula of  $A_2B_2O_7$  [21–27]; however, materials are often tested by different protocols with different environmental variables (such as temperature, surface area/leachate volume ratios, leachates), sample geometries (e.g., powders vs. densified pellets) and microstructure. The experimentally determined elemental release rates may vary by several orders of magnitudes for the same materials, making various experimental data not comparable and creating challenges to achieve mechanistic understanding and predict materials' performance. Therefore, it is essential to test materials having similar microstructure and geometry and to use the same protocol with controlled environmental variables.

However, chemical durability testing typically requires a great amount of time and effort in order to assess materials' potentials for waste form applications. It is a huge scientific challenge and technically infeasible to perform leaching experiments across hundreds of thousands of possibilities of the pyrochlore and fluorite structure types given their composition complexity with a large design window of compositions. A computational-based approach by considering key materials and structural characteristics governing materials' chemical durability is highly desirable to accelerate materials design and discovery in order to optimize materials compositions and performance. Recently, machine learning has become an indispensable technique in accelerating the design of new materials and evaluating materials properties for nuclear waste form development with respect to the traditional trial and error approaches [28]. Data-driven machine learning approaches have become the fourth paradigm along the material science development path, following the pure empirical (via observations), formula and laws, and computational science paradigms [29]. In certain circumstances, data-driven approaches such as machine learning would be appropriate

and more efficient [30]. This is of particular interest for material properties difficult to be measured experimentally or affected by complicated phenomena where no simple governing equations are available. For example, in the scenario of diverse crystal chemistry and composition complexity of the  $A_2B_2O_7$  structural types and unrevealed corrosion mechanisms and kinetics, data-driven processes are necessary to predict materials performance and accelerate materials design and discovery.

Machine learning approaches have been applied to the design of ceramic waste form materials and the prediction of properties (e.g., radiation tolerance and chemical durability) by building machine learning models, analyzing experimental data, and predicting their performance. For example, on waste form design, artificial neural network (ANN) has been employed to predict new apatite and hollandite compositions out of a wide range of possibilities. In particular, a crystal chemistry approach has been applied using ANN to explore the compositional space of promising hollandite and apatite compositions for iodine and Cs incorporations [31,32]. Illington et al. used machine learning algorithms to predict the static and dynamic leaching behavior of borosilicate glass matrix [33] and assessed the performance with a large set of data available in the literature [34]. However, the datasets may be obtained from multiple sources, and non-uniform or inconsistent data acquisition processes raise questions concerning the quality and accuracy of these datasets. Lumpkin et al. [35] studied the correlation between radiation tolerance and a series of parameters such as electronegativity difference and disorder energy. Regression models were generated based on experimental data, and statistical models were derived to predict the critical temperature in radiation, an essential performance indicator for pyrochlore and fluorite structure-based waste form materials. These models generally achieve  $> 90\% R^2$  and demonstrate exceptional predictability despite the fact that the linear regression model is constrained when the correlations among parameters become more complex. Krishnan et al. [36] utilized multiple machine learning models, including linear, support vector machine, artificial neural network, etc., to predict silicate glass leaching kinetics. Although simple models such as linear regression may not lead to a reasonable predictive capability, additional physics-based constraints might greatly improve the performance. Zhang et al. established a data-driven model with a large dataset to evaluate and predict the chemical durability of oxide glasses [37]. The authors trained and tested various machine learning models, including random forest (RF), artificial neural network (ANN), and k-nearest neighbors (KNN). Among different models, the ANN model has the highest accuracy in modeling the weight loss of the glass under different chemical conditions, while RF has higher accuracy in determining the glass surface appearance change. RF and ANN models were not considered here because of the limited database in the current work. As a rule of thumb, with a limited database, traditional statistical models such as linear regression or logistic regression, as well as support vector machine performing kernel trick, are more suitable than more advanced models such as ANN or RF. Tree-based models are more appropriate to be applied on the middle-size dataset, while ANN generally has more applications when large datasets are available.

In this work, machine learning models are developed based on a set of limited data of chemical durability of 30 compositions in a general formula of  $A_2B_2O_7$  ranging from single component rare earth (Sm to Yb) titanates, zirconates, and stannates to multiple component solid solutions and high entropy ceramics. These 30 compositions of pyrochlorides and fluorite structural types were synthesized by spark plasma sintering (SPS) with similar microstructure and grain size. The as-purchased rare earth oxides and  $TiO_2$  precursor powders were first pre-reacted via high-energy ball milling following by the SPS consolidation at different temperatures, duration, and pressure. The detailed materials synthesis and characterization can be found in our previous publications [38]. All the sample pellets were polished following the same procedure before conducting semi-dynamic leaching tests. Chemical durability and elemental release were determined at identical experimental conditions using the same semi-dynamic leaching testing protocol, providing a

scientific basis for evaluating materials performance and understanding their dissolution kinetics and elemental release behavior [26,27,35,39–44]. The details of the semi-dynamic leaching tests can be referred to in previous literature [38,45]. The semi-dynamic leaching tests were conducted in a closed cap PTFE vial in a digital controlled oven at a temperature of 90 °C. The surface area/volume ratio for the leaching test is 5.0, with the leachate being exchanged per three hours for the first 24 h and per day for the rest 14 days. The leachates were then analyzed through an ICP-MS in order to quantify the elemental concentration. The experimentally determined short-term and long-term leaching rates of the A-site cations are tabulated in Table 1, along with the materials and structural parameters used as the input parameters and structural descriptors to the machine-learning models. Despite a small set of data across 30 compositions, the structural characteristics can be well correlated to cation substitution at both A- and B-sites within a confined structure. Therefore, machining learning models can be derived with a high fidelity based on both short-term and long-term leaching rates of each composition, enabling the prediction of materials' chemical durability of hundreds of thousands of compositions of the  $\text{A}_2\text{B}_2\text{O}_7$  structure types.

To realize this, the small set of leaching rate data were randomly divided into training and testing datasets, which then are used to train and test machine learning models. The performance of the models is evaluated based on the predictability on the testing datasets not used for training the models. The results indicate that the models have reasonable predictability, demonstrating the great potential of using machine learning approaches to estimate the leaching rates of the  $\text{A}_2\text{B}_2\text{O}_7$  structure types in order to speed up the materials discovery and performance prediction. Key parameters as summarized in Table 1, such as ionic radii size difference  $r_{\text{A}-\text{B}}$ , ionic potential difference  $E_{\text{B}-\text{A}}$ , electronegativity difference  $\chi_{\text{B}-\text{A}}$ , and lattice parameter  $a$ , are considered as the structural descriptors for the machining learning model development and key factors governing materials' chemical durability and

leachability are revealed based on the minimized root mean square errors. These four features were selected because they are very fundamental and important structural and energetic properties of the material, which determines the performance of the materials in many aspects, such as the amorphization resistance [46]. In a recent paper identifying the critical features for predicting the glass-forming ability [47], the authors used an approach to use RF models for feature selection. The new model was reported to have a 12% increase in prediction accuracy. In this work, additional parameters such as the bond-valence sum of lanthanides and oxygen positional parameters were also considered, which do not produce more accurate prediction. Due to the limited datasets, only the aforementioned 4 features were selected to train the model to avoid possible overfitting, as more features generally require larger datasets. The derived linear regression and Kernel ridge regression (KRR) models can be used to predict the short-term and long-term leaching rates for other components, making it feasible to identify the most suitable compositions with optimized materials chemical durability as required for waste form materials for long-term nuclear waste management.

## 2. Computational methods

Among different machine learning algorithms, linear regression is widely used to combine different factors with various weights, which is simple to implement. The general form of multivariate linear regression with  $n$  variables is shown in formula (1), where  $X_i$  and  $Y$  are the corresponding input variables and the output variable,  $\beta_0$  is known as the intercept, and  $\beta_i$  are the fitting coefficients. The regression algorithm normally fits the linear model with coefficients that can minimize the residual sum of squares between the predicted values and the observations. IN this work, the linear regression models for both short-term and long-term leaching data (Table 1) of the  $\text{A}_2\text{B}_2\text{O}_7$  structure type were generated with a Scikit-learn machine learning package [48].

**Table 1**

The summary of leaching behavior and key parameters of various families of Pyrochlores obtained through semi-dynamic leaching following an ASTM C1308 protocol and dynamic leaching testing. The data for Ti and Zr families was reported in [38,45] while the other data are under preparation for publication.

Element	A	rA - rB	Ionic (B - A)	Electronegativity (B - A)	lattice parameter	short term rate (mmol/m <sup>2</sup> /hr)	long term rate (mmol/m <sup>2</sup> /d)
Ti	Sm	0.474	3.8312	0.49	10.23	0.240	0.015
	Gd	0.448	3.7625	0.46	10.19	0.193	0.011
	Dy	0.422	3.6904	0.43	10.08	0.125	0.006
	Er	0.399	3.6235	0.41	10.07	0.114	0.005
	Yb	0.380	3.5659	0.39	10.04	0.092	0.005
	Sm,Yb	0.427	3.6986	0.44	10.14	0.163	0.008
	Sm,Yb,Gd	0.434	3.7206	0.45	10.15	0.186	0.010
	Sm,Yb,Gd,Er	0.425	3.6956	0.44	10.13	0.160	0.008
	Sm,Yb,Gd,Er,Dy	0.424	3.6946	0.44	10.13	0.149	0.008
	Nd	0.389	2.8550	0.10	10.67	0.174	0.009
Zr	Sm	0.359	2.7796	0.08	10.59	0.167	0.004
	Gd	0.333	2.7109	0.05	10.53	0.043	0.004
	Dy	0.307	2.6388	0.02	9.93	0.042	0.003
	Er	0.284	2.5719	0.00	9.79	0.010	0.000
	Yb	0.265	2.5143	-0.02	9.66	0.007	0.000
Sn	La	0.470	3.2110	0.99	10.71	0.393	0.012
	Nd	0.419	3.0920	0.66	10.58	0.380	0.010
	Sm	0.389	3.0166	0.64	10.52	0.285	0.010
	Gd	0.363	2.9479	0.61	10.46	0.300	0.011
	Dy	0.337	2.8758	0.58	10.40	0.323	0.006
U doping	Er	0.314	2.8089	0.56	10.36	0.255	0.004
	Yb	0.380	3.5704	0.38	10.05	0.150	0.005
	Sm, Yb	0.435	3.6674	0.42	10.14	0.201	0.010
	Sm, Yb, Gd	0.442	3.6878	0.43	10.18	0.236	0.009
	Sm, Yb, Gd, Er	0.434	3.6647	0.42	10.16	0.224	0.009
Zr	Sm, Yb, Gd, Er, Dy	0.423	3.6928	0.43	10.08	0.149	0.013
	(Sm, Yb) <sub>2</sub> Zr <sub>2</sub> O <sub>7</sub>	0.312	2.6469	0.03	10.13	0.120	0.002
	(Sm, Yb, Gd) <sub>2</sub> Zr <sub>2</sub> O <sub>7</sub>	0.319	2.6683	0.04	10.26	0.148	0.003
	(Sm, Yb, Gd, Er) <sub>2</sub> Zr <sub>2</sub> O <sub>7</sub>	0.310	2.6442	0.03	10.12	0.140	0.002
	(Sm, Yb, Gd, Er, Dy) <sub>2</sub> Zr <sub>2</sub> O <sub>7</sub>	0.309	2.6431	0.03	10.04	0.110	0.002

$$Y = \beta_0 + \sum_{i=1}^n X_i \beta_i \quad (1)$$

When linear regression is not adequate in capturing complex relationships, other machine learning methods can be used to model nonlinear relations between the output and input. Kernel ridge regression (KRR) model is one of these machine learning models, which has been widely applied in the literature to train and predict various properties of materials. For example, Rupp et al. [49] used the KRR model to predict the atomization energies of organic molecules based on 2 parameters. Stuke et al. [50] studied the molecular orbital energy using data from several large databases to explore the energy of molecules and accelerate the material analysis process. Pilania *et al.* used KRR models to determine the amorphization behavior of pyrochlores [46]. Giguere *et al.* proposed a new kernel to be used in KRR, which generalized many of the common kernels. These authors demonstrated that KRR is better than a support vector regression (SVR) model in the sense of easy tuning and better predicting performance. The principle of KRR is to include regularization to the coefficients through introducing penalty terms [51]. The KRR estimator is written in Eq. (2) [52], where  $x_i$  and  $y_i$  are pairwise covariates and the response variable,  $\lambda$  is the regularization parameter, and  $H$  stands for the kernel Hilbert space. In the current work, five-fold cross-validation with a validation set size being 20% was used to protect against overfitting and to tune the hyperparameter that governs the KRR kernel, including regularization strength, kernel type, and gamma that is the inverse of the standard deviation of a Gaussian function. Grid search was applied to find the best combination of the hyperparameters. For both linear and KRR models, all features were pre-standardized such that each feature centered at the corresponding mean with a standard deviation. The standardization is a general and common approach for data preprocessing since many kernels and regularizers require predictors to have variance in similar magnitude.

$$\hat{f} := \operatorname{argmin} \left\{ \frac{1}{N} \sum_{i=1}^N (f(x_i) - y_i)^2 + \lambda \|f\|_H^2 \right\} \quad (2)$$

## 2.1. Training and testing datasets

The overall dataset contained 30 data points, which were obtained through semi-dynamic leaching experiments (see Table 1) [43,44] and was divided into two subsequent datasets. The training dataset consists of 24 data points, occupying 80% of the total dataset. On the other hand, the testing dataset consists of 6 data points (20% of the total dataset). All the leaching rates used in machine learning were derived from the lab-scale 14-day semi-dynamic leaching tests, in which we strictly followed the procedure reported in our previous research [38,45]. Both the short-term and long-term leaching rates can be found in Table 1. The input data considered here include ionic radius difference  $r_{A-B}$ , ionic potential difference  $E_{B-A}$ , electronegativity difference  $\chi_{B-A}$ , and lattice parameter  $a$ . The response variables are short-term leaching rates and long-term leaching rates, with units being  $\text{mmol}\cdot\text{m}^{-2}\cdot\text{hour}^{-1}$  and  $\text{mmol}\cdot\text{m}^{-2}\text{day}^{-1}$ , respectively. The lattice parameters were derived from the XRD peak refinements. The other factors, including ionic radius, electronegativity difference, and ionic potential difference, were found from the literature [53–56].

## 2.2. Feature selection and hyperparameter tuning

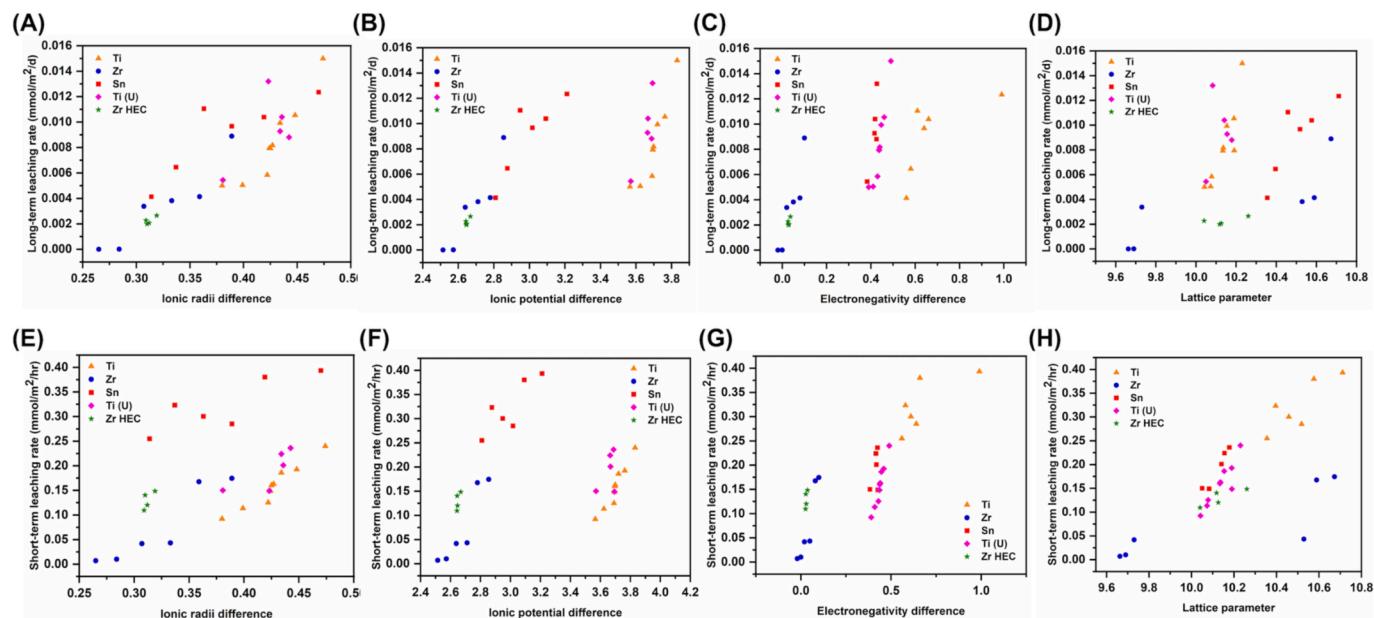
Four key materials parameters were considered as the most important factors governing the leaching behavior of the  $\text{A}_2\text{B}_2\text{O}_7$  structure types. However, the role played by each factor may have different significance. In order to avoid potential overfitting, feature selection from these properties was performed to progressively filter out the factors that are with the least significance. The feature selection for the linear model was based on a leave-one-out version of the cross-validation, where the training dataset was further divided into a smaller training

set and a validation set of size one. For a training dataset with size  $N$ , there are  $N$  different ways for the division, and thus the cross-validation error is given in equation (3), where  $e_n$  denotes the error on each validation set and can be calculated using squared differences between the predicted value and the observed value. For the RKK model, 5-fold cross-validation was used. The process of hyperparameter tuning involves the selection of a set of optimal hyperparameters for a specific learning problem. For the RKK model, grid search, an exhaustive search over designated parameter ranges, was adopted. Specifically, the hyperparameter  $\alpha$  and  $\gamma$  were tuned, with the former being the regularization strength that reduces variance and avoids overfitting and the latter being a parameter for the RBF kernel that defines the influence of a single training data point. The searching range for  $\alpha$  is from 1E to 3 to 1, with a linear increment of 1E-3, while the searching range for  $\gamma$  is from  $2^{-5}$  to  $2^5$  with a proportional increment of 2.

$$E_{cv} = \frac{1}{N} \sum_{n=1}^N e_n \quad (3)$$

## 3. Results and discussion

We first show the correlation of the long-term and short-term leaching rates of all the samples against each individual feature, as shown in Fig. 1. Although simple correlations can be obtained between the leaching rate and each feature, no single feature is adequate to individually predict the long-term and short-term leaching behavior. For example, in Fig. 1A, the long-term leaching rate shows an excellent correlation with ionic radius difference ( $r_{A-B}$ ), but the short-term leaching rate shown in Fig. 1E suggests that Sn-based pyrochlore has a different response than the other compositions. Both the long-term and short-term leaching rates increase when the ionic radius difference increases, as indicated in Fig. 1A and E. It has been suggested that chemical durability is related to both the metal–oxygen bonding energy and metal–metal repulsion [57]. Since the A–O distance in the  $\text{A}_2\text{B}_2\text{O}_7$  structure decreases with the incorporation of smaller A-site cations (from La to Yb), an increase in the A–O bonding energy is anticipated [58]. The stability of the  $\text{A}_2\text{B}_2\text{O}_7$  structure-types strongly depends on the ionic radius difference between the A-site and B-site elements. As the cations at A and B sites become more similar in size, the structure is changed from the ordered pyrochlore superstructure to an anion deficient fluorite structure by disordering the A-site and B-site cations and the oxygen sublattice. The order to disorder transition of its crystalline structure and the oxygen displacement from the nearest neighbor site leads to enhancement of materials' radiation tolerance and may also affect its chemical durability [59]. The A–O bonding length gradually decreases as the bonding strength increases with a small A-site ionic radius, which can be confirmed by the previous research [60]. However, the difference of Sn-based pyrochlore structure with respect to other compositions shown in Fig. 1E suggests a significant influence of order-disorder transition on materials' chemical durability due to the potentially strong impact of B-site elements. Both long-term and short-term leaching rates are positively correlated with the feature ionic potential difference  $E_{B-A}$  (Fig. 1B and F), and the slope of the leaching data of each family is close. However, there is an obvious separation between the Ti-based family and the other families of the  $\text{A}_2\text{B}_2\text{O}_7$  structural types. The same observation can be found for the electronegativity difference, where the Zr-based family is distinct from the other families (Fig. 1C and G). Lastly, for the feature lattice parameter, short-term leaching behavior is reasonably correlated with the lattice parameter shown in Fig. 1H, but the long-term behavior does not show a very meaningful correlation. In the previous research conducted by Brik *et al.* [60], the lattice parameter can be described by ionic radius and the difference of electronegativity of the cations. Lattice parameter has a positive correlation to ionic radii of both A-site and B-site cations, while a negative correlation can be observed for electronegativity as suggested by the model in a previous study [60]. Therefore, the study herein is generally



**Fig. 1.** The correlations between the long-term leaching rate (A-D) and short-term leaching rate (E-H) with 4 material parameters/structural features, including ionic radius difference, ionic potential difference, electronegativity difference, and lattice parameter. No single feature is solely adequate to predict long-term and short-term leaching behavior.

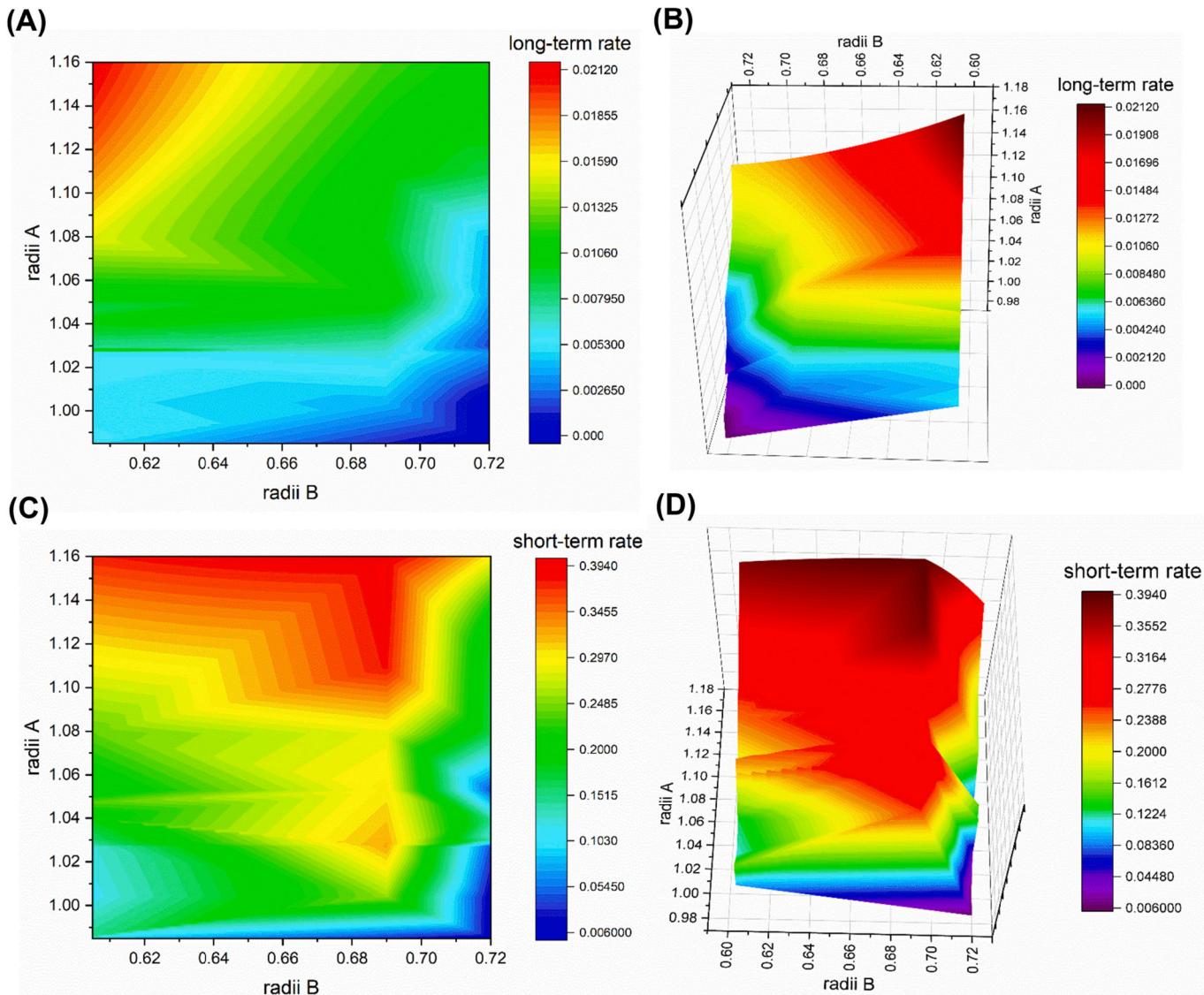
consistent with the discovery in which materials' chemical durability is strongly correlated to the bonding strength. Specifically, the metal–oxygen bonds (A-O bonds) in the coordinated polyhedron become stronger by reducing the A-site ionic radius and thus greater ionic potential, resulting in lower elemental release rates [55]. On the other hand, electronegativity reflects the bonding nature of atomic pairs in oxides, which can also be related to bonding types and bonding length. The electronegativity of elements can be characterized by ionization potential and electron affinity. The ionization potential is fairly related to the covalent radius, from which it is reasonable to infer that electronegativity is strongly associated with chemical durability [61]. Therefore, the higher of the ionic potential and electronegativity of lanthanides with reduced ionic radii from La to Yb, the stronger will be the covalent bonds between A-site element and oxygen, which results in lower leaching rates for the short term and long term as well. In addition to ionic potential and electronegativity difference, a systematic increase in the electron density at the A-site nuclei is anticipated based on the variation in bonding length and the reduction in atomic volume. However, the long-term leaching behavior is not fully correlated to the nature of the bonding due to complications resulting from different surface alteration behavior. Particularly, titanate pyrochlore is more prone to form an amorphous surface passivation film, greatly reducing the long-term leaching rates compared to zirconates [38,45]. Therefore, the strong covalent bonding within the  $A_2Zr_2O_7$  may not necessarily result in a lower long-term leaching rate when compared with isostructural titanate compositions, and this is consistent with a previous result that the long term release rates of  $Gd_2Zr_2O_7$  and  $Gd_2Ti_2O_7$  are comparable [62]. Overall, all features demonstrate some predictability in estimating leaching behaviors of the  $A_2B_2O_7$  structure-types, but no single feature is sufficient to individually make reliable predictions across a wide range of  $A_2B_2O_7$ , showing extreme composition complexity.

Apart from the single variable regression, regressions with 2 variables were also derived to determine the interrelationship between the response leaching rate and multiple dependent input variables. Fig. 2 shows the contour mapping with the response variable being the long-term and short-term leaching rates and predictor variables being ionic radii of the A cation and B cation. Except for the original 30 data points, the other data points are obtained through a scattered interpolation tool in MATLAB®, which generates an interpolating function that can be

efficiently used to query values of the response variable. Several observations can be made from Fig. 2. A larger value of radius A and radius B generally leads to a higher leaching rate for long-term leaching. This can be attributed to the smaller bonding energy in the larger-sized lanthanide-oxygen bond (e.g., La-O) over those Yb-O and Er-O bonds. The metal–metal repulsion also decreases between A- and B-site cations from  $La_2B_2O_7$  to  $Yb_2B_2O_7$  [58], reducing the elemental release rate. A small discrepancy can be observed in Fig. 2C where radii of A-site cations are in the range of 1.01 to 1.05 and radii of B-site cations are in the range of 0.67 to 0.7, corresponding to the deviation shown in Fig. 1E where the Sn-based family shows a different response than the other families.

### 3.1. Results of multivariate linear regression and machine learning

More complicated correlations between predictors and response variables can be depicted using multivariate linear regressions. In order to determine a combination of the features that has the best performance, the approach described in [46] was applied, where the root mean square error (RMSE) for the predictions on training and testing sets were used to determine the best combination of features. The performance of each feature combination trained on the same randomly-split training dataset and evaluated on the training dataset and the validation dataset is displayed in Fig. 3. Adding more features to the combination reduces the RMSE in general. However, attention needs to be paid to avoid overfitting as too many features may possibly lead to reduced reliability of the prediction power. Fig. 3A shows the results of the feature selection for the long-term linear model. For models with a single feature, the model with ionic radius difference  $r_{A-B}$  has the best performance, suggesting that the ionic radius difference  $r_{A-B}$  has the best predictability in determining the long-term leaching rate, which is consistent with the observations in Fig. 1A and 1E. On the other hand, the other three 1-D models with a higher RMSE do not show good performance when these features are used alone. Considering the group of 2D models, it can be observed that the first three models containing ionic radius difference have better performance than the other three models. This further confirms that the feature ionic radius difference is the most significant one dominating the long-term leaching behavior. This is due to the fact that ionic radius has a significant impact on the



**Fig. 2.** 2D and 3D contour maps of the long-term and short-term leaching rates vs. ionic radii A and ionic radii B. All contour maps were constructed from experimentally-measured leaching data and the interpolated data generated using a scattered interpolation tool in MATLAB.

metal–oxygen bonding energy, consistent with the previous research conducted by Chakoumakos [57]. Although lattice parameter has a strong correlation to the ionic radius difference, other factors, including electronegativity and ionic potential, also play important roles in governing the changes in lattice parameter [60]. For the group of the 3D models, it can be observed that the feature combination of ionic radius difference  $r_{A-B}$ , ionic potential difference  $E_{B-A}$ , and electronegativity difference  $\chi_{B-A}$ , has the best performance with the lowest RMSE. On the contrary, the other three models with lattice parameter have elevated RMSE values. Further increasing the number of features leads to a limited effect on the diminishment in the RMSE. Finally, the lowest RMSE for all combinations of features is achieved with all four features. Thus, all 4 features with ionic radius size difference  $r_{A-B}$ , ionic potential difference  $E_{B-A}$ , and electronegativity difference  $\chi_{B-A}$ , and  $\alpha$ , will be used to train the model. The overall performance of the KRR model is similar to that of the linear model, which indicates that the correlation among predictors and the response variables is not complicated, mostly linear in nature. The RMSE results from the KRR model are slightly different from the linear model. For example, in the 1D group, the parameter ionic potential difference  $E_{B-A}$  leads to the lowest RMSE, deviating from the linear model. This suggests that different regression

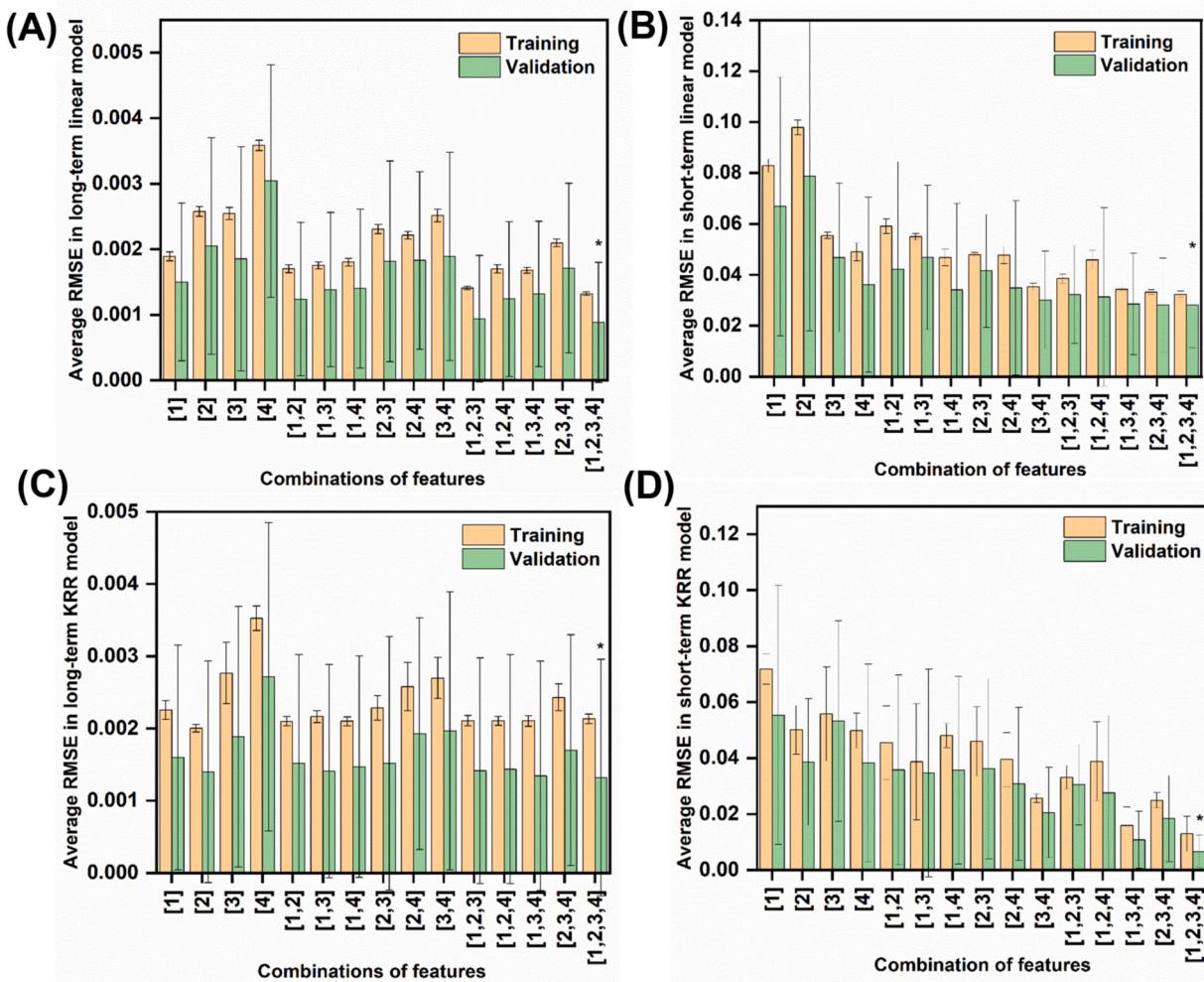
models have different predictabilities and thus different performance with the same predictors. The results are similar for short-term leaching rates. The lowest RMSE is achieved with the combination of all four parameters.

The linear regression models for long-term and short-term leaching rates are shown in formula (4–5). From the explicit models below, for both long-term and short-term leaching, the leaching rate is positively associated with ionic radius and electronegativity differences, while negatively related with ionic potential difference between A- and B-site cations. The R-squared value for long-term and short-term leaching using linear regression is 0.867 and 0.893, respectively, indicating that the models are appropriate and suitable to explain the training data.

$$K_l = -0.0200 + 0.0565(r_{A-B}) - 0.0021(E_{B-A}) + 0.0036(\chi_{B-A}) - 0.0010(\alpha) \quad (4)$$

$$K_s = -0.5985 + 0.4461(r_{A-B}) - 0.0801(E_{B-A}) + 0.2915(\chi_{B-A}) + 0.0742(\alpha) \quad (5)$$

The performance of both regression models was subsequently evaluated by predicting leaching rates based on the input of the testing set. Fig. 4A shows the plot comparing the true data obtained from

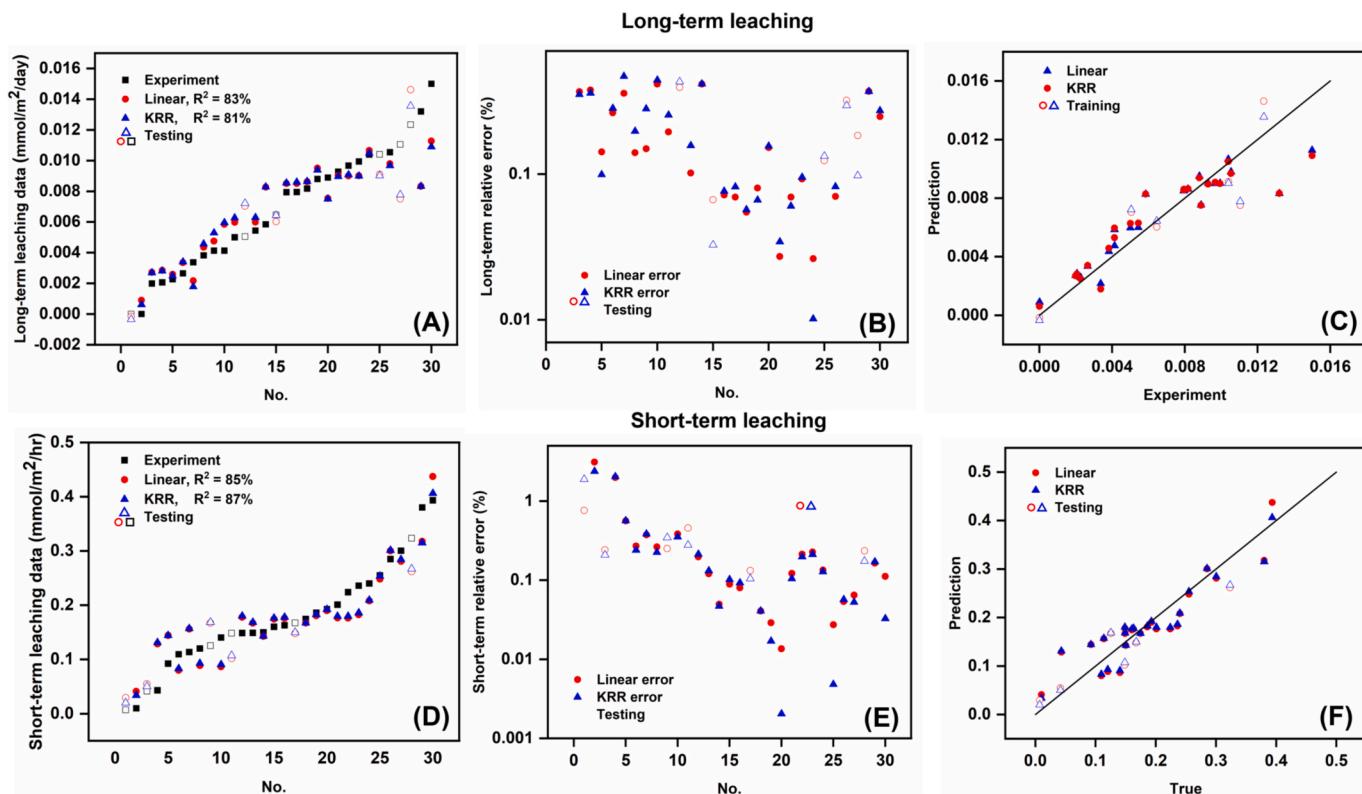


**Fig. 3.** Average RMSE obtained from the linear regression and KRR models trained with various combinations of features to the randomly-split training and validation dataset. Specifically, RMSE was evaluated based on (A) a long-term linear model, (B) a short-term linear model, (C) a long-term KRR model, and (D) a short-term KRR model when predicting the leaching rates. Both training error and validation errors are shown, and the combination of all 4 features leads to the lowest RMSE.

experiments, along with the predicted values from the linear regression and KRR models for long-term leaching. The data from the training set are shown in solid symbols, while the data from the testing set are shown in the corresponding open-circle symbols. All of the data are arranged in an ascending order of the values. It can be seen that the predicted values from both models match well with the experimental values. The linear model achieves an R-squared value of 83%, and the KRR model achieves 81%, which suggests that the majority of the variance in the long-term leaching rate can be explained by these four predictors. KRR displays a slightly lower R-squared value than linear regression. However, this does not necessarily mean that KRR has worse performance than the linear model all the time, since the dataset split is random, and the linear model may have a better response in another dataset split. Fig. 4B shows the relative error of each model with respect to the experimental data. The majority of the error is in the range of 1% to 10%, indicating that both models have excellent predictability. Fig. 4C displays the parity plot, along with a reference line ( $y = x$ ) to facilitate visualization. The more data points close to the reference line, the better the prediction is. The majority of the data points are either on or close to the reference line, which clearly indicates that both models can reasonably predict the long-term leaching rate. No obvious differences can be found between the training set and testing set, suggesting no overfitting in both models. Similar to long-term leaching experiments, the short-term leaching results are shown in Fig. 4D, where one can find that both models can

make accurate predictions on the leaching rate. The linear model achieves an R-squared of 85%, and the KRR model achieves 87%. This further confirms that the linear regression and machine learning models are capable of predicting the leaching behavior of  $\text{A}_2\text{B}_2\text{O}_7$  structural types and have comparable performance. Fig. 4E shows the relative error in the short-term leaching rate. The majority of the error is in the range of 1% to 15%. Several outliers show higher errors mainly because the experimental value of these data points is very small, leading to higher relative error. From the parity plot shown in Fig. 4F, one can see that the data points are very close to the reference line, indicating all the predicted values from linear and KRR models are very close to their corresponding experimental values. As in the case of long-term leaching, no obvious distinction is observed between the training set and the testing set, suggesting no overfitting in the models.

The statistical and machine learning approaches are applied to both long-term and short-term leaching scenarios and achieve reasonable predictability. This demonstrates the robustness and validity of the approaches, which can be used for the exploration of other properties of pyrochlores. Considering only four parameters used, it would be beneficial and simple in determining the leaching performance of other unexplored  $\text{A}_2\text{B}_2\text{O}_7$ , including multicomponent solid solution oxides across a wide range of design space, as long as the data of the aforementioned four input variables are available.



**Fig. 4.** Comparison between the experimental data and the predicted values with the linear regression model and KRR model on (A) long-term leaching data and (D) short-term leaching data. The plot showing the relative error of the predicted value comparing to the experimental value on (B) long-term leaching data and (E) short-term leaching data. The parity plot showing that the predicted value match very well with the experimental data on (C) long-term leaching data and (F) short-term leaching data. For all plots, the training datasets are displayed with the solid symbols, while the testing datasets are displayed with the corresponding non-solid symbols.

#### 4. Conclusions

A systematic machine learning investigation has been carried out to search for the most significant parameters of the  $\text{A}_2\text{B}_2\text{O}_7$  structural types that govern their corrosion performance and materials' chemical durability based on a limited set of elemental release rates obtained under identical leaching testing. Both long-term and short-term leaching rates are evaluated, and their corresponding correlations with the key structural and compositional parameters are analyzed. Four parameters are determined to be the most critical, governing the leaching rates with model validation, which include ionic radius difference, ionic potential difference, electronegativity difference between A-site and B-site cations, and their lattice parameters. The experimental data are further used to construct two machine learning models based on randomly split training and testing datasets. All four features are necessary for predicting the short-term and long-term leaching rates, with ionic radius difference, ionic potential difference, and electronegativity difference between A-site and B-site cations demonstrate a positive correlation to the elemental release rates. Generally, the modeling can well predict the trend of the elemental release rates experimentally determined from  $\text{Sm}_2\text{Ti}_2\text{O}_7$  with the highest elemental rate ( $0.015 \text{ mg/m}^2/\text{d}$ ) and to  $\text{Yb}_2\text{Zr}_2\text{O}_7$  with the lowest rate ( $0.00001 \text{ mg/m}^2/\text{d}$ ). The test datasets indicate that the machine learning models can achieve an overall  $R^2$  higher than 80%, suggesting excellent predictabilities, and the relative errors between the experimental and the predicted values are approximately 10%. The explicit multivariate linear model derived can be utilized to quickly predict the leaching rates of the  $\text{A}_2\text{B}_2\text{O}_7$  structural types with identified key materials and structural parameters as the modeling descriptors. These results highlight that machine learning techniques can be applied to predict material properties (e.g., chemical durability in this work) with reasonable accuracy. This approach

provides a practical tool to reduce the experimental efforts and predict the performance of materials across a wide range of materials with extreme composition complexity in order to accelerate materials discovery with a large design window.

#### CRediT authorship contribution statement

**Bowen Gong:** Conceptualization, Investigation, Visualization, Writing – original draft, Writing - review & editing. **Kun Yang:** Conceptualization, Investigation, Visualization, Writing – original draft, Writing - review & editing. **James A. Lian:** Conceptualization, Investigation, Visualization, Writing - review & editing. **Jianwei Wang:** Conceptualization, Writing - review & editing.

#### Declaration of Competing Interest

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

#### Acknowledgements

This work was supported as part of the Center for Performance and Design of Nuclear Waste Forms and Containers (WastePD), an Energy Frontier Research Center (EFRC) funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0016584.

#### Data availability

The data that support the findings of this study are available from the

corresponding author upon reasonable request.

## References

- [1] R.C. Ewing, W.J. Weber, J. Lian, *J. Appl. Phys.* 95 (2004) 5949–5971.
- [2] J. Lian, K. Helean, B. Kennedy, L. Wang, A. Navrotsky, R. Ewing, *J. Phys. Chem. B* 110 (2006) 2343–2350.
- [3] F. Brauer, R. Stromatt, J. Ludwick, F. Roberts, W. Lyon, *J. Inorg. Nucl. Chem.* 12 (1960) 234–235.
- [4] S. Gin, A. Abdelouas, L.J. Criscenti, W.L. Ebert, K. Ferrand, T. Geisler, M. T. Harrison, Y. Inagaki, S. Mitsui, K.T. Mueller, *Mater. Today* 16 (2013) 243–248.
- [5] E. Vernaz, J. Dussossoy, *Appl. Geochem.* 7 (1992) 13–22.
- [6] C. Jantzen, M. Plodinec, *J. Non-Cryst. Solids* 67 (1984) 207–223.
- [7] P. Morgan, D. Clarke, C. Jantzen, A. Barker, *J. Am. Ceram. Soc.* 64 (1981) 249–258.
- [8] W. Weber, R. Ewing, C. Catlow, T.D. De La Rubia, L. Hobbs, C. Kinoshita, A. Motta, M. Nastasi, E. Salje, *E. Vance* 13 (1998) 1434–1484.
- [9] R. Roy, E. Vance, J. Alamo, *Mater. Res. Bull.* 17 (1982) 585–589.
- [10] E.R. Merz, C.E. Walter (Eds.), *Disposal of Weapon Plutonium*, Springer Netherlands, Dordrecht, 1996.
- [11] R.C. Ewing, *Proc. Natl. Acad. Sci.* 96 (1999) 3432–3439.
- [12] S.V. Stefanovsky, S.V. Yudintsev, R. Gieré, G.R. Lumpkin, *Geological Society, London, Special Publications* 236 (2004) 37–63.
- [13] K. Yang, K. Bryce, W. Zhu, D. Zhao, J. Lian, *J. Eur. Ceram. Soc.* 41 (2021) 2870–2882.
- [14] D. Strachan, R. Turcotte, B. Barnes, *Nucl. Technol.* 56 (1982) 306–312.
- [15] D. Strachan, *Nucl. Chem. Waste Manage.* 4 (1983) 177–188.
- [16] S.-B. Xing, I.L. Pegg, *MRS Online Proc. Library* 333 (1993) 557–564.
- [17] C.M. Jantzen, N.E. Bibler, *Ceram. Trans.* 176 (2006) 141.
- [18] ASTM C1220-17. Standard Test Method for Static Leaching of Monolithic Waste Forms for Disposal of Radioactive Waste, ASTM International, West Conshohocken, PA, 2017.
- [19] ASTM C1308-21. Standard Test Method for Accelerated Leach Test for Measuring Contaminant Releases From Solidified Waste, ASTM International, West Conshohocken, PA, 2021.
- [20] G.S. Frankel, J.D. Vienna, J. Lian, X. Guo, S. Gin, S.H. Kim, J. Du, J.V. Ryan, J. Wang, W. Windl, C.D. Taylor, J.R. Scully, *Ceram. Rev.*, in press (2021).
- [21] L. Peng, K. Zhang, D. Yin, J. Wu, S. He, H. He, *Ceram. Int.* 42 (2016) 18907–18913.
- [22] B.D. Begg, N.J. Hess, W.J. Weber, R. Devanathan, J.P. Icenhower, S. Thevuthasan, B.P. McGrail, *J. Nucl. Mater.* 288 (2001) 208–216.
- [23] J.P. Icenhower, B.P. McGrail, H.T. Schaeff, E.A. Rodriguez, *MRS Online Proc. Library Arch.* 608 (1999).
- [24] K. Liu, K. Zhang, T. Deng, J. Zeng, B. Luo, H. Zhang, *Ceram. Int.* 47 (2021) 13363–13373.
- [25] J.P. Icenhower, D.M. Strachan, B.P. McGrail, R.D. Scheele, E.A. Rodriguez, J. L. Steele, V.L. Legore, *Am. Mineral.* 91 (2006) 39–53.
- [26] H. Xu, Y. Wang, P. Zhao, W.L. Bourcier, R. Van Konynenburg, H.F. Shaw, *Environ. Sci. Technol.* 38 (2004) 1480–1486.
- [27] D.M. Strachan, R.D. Scheele, E.C. Buck, J.P. Icenhower, A.E. Kozelisky, R.L. Sell, R. J. Elovich, W.C. Buchmiller, *J. Nucl. Mater.* 345 (2005) 109–135.
- [28] J. Wei, X. Chu, X.Y. Sun, K. Xu, H.X. Deng, J. Chen, Z. Wei, M. Lei, *InfoMat* 1 (2019) 338–358.
- [29] A. Agrawal, A. Choudhary, *APL Mater.* 4 (2016) 053208.
- [30] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *NPJ Comput. Mater.* 3 (2017) 1–13.
- [31] J. Wang, *Front. Earth Sci.* 3 (2015) 20.
- [32] D.B. Ghosh, B.B. Karki, J. Wang, *J. Nucl. Mater.* 530 (2020) 151957.
- [33] J.N. Lillington, T.L. Goût, M.T. Harrison, I. Farnan, *J. Non-Cryst. Solids* 533 (2020) 119852.
- [34] J.N. Lillington, T.L. Goût, M.T. Harrison, I. Farnan, *J. Non-Cryst. Solids* 546 (2020) 120276.
- [35] G.R. Lumpkin, M. Pruneda, S. Rios, K.L. Smith, K. Trachenko, K.R. Whittle, N. J. Zaluzec, *J. Solid State Chem.* 180 (2007) 1512–1518.
- [36] N.A. Krishnan, S. Mangalathu, M.M. Smedskjaer, A. Tandia, H. Burton, M. Bauchy, *J. Non-Cryst. Solids* 487 (2018) 37–45.
- [37] Y. Zhang, A. Li, B. Deng, K.K. Hughes, *NPJ Mater. Degrad.* 4 (2020) 1–11.
- [38] K. Yang, P. Lei, T. Yao, B. Gong, Y. Wang, M. Li, J. Wang, J. Lian, *Corros. Sci.* 185 (2021) 109394.
- [39] J. Lian, J. Chen, L. Wang, R.C. Ewing, J.M. Farmer, L.A. Boatner, K. Helean, *Phys. Rev. B* 68 (2003) 134107.
- [40] J. Zhang, J. Lian, A.F. Fuentes, F. Zhang, M. Lang, F. Lu, R.C. Ewing, *Appl. Phys. Lett.* 94 (2009) 243110.
- [41] S. Lutique, D. Staicu, R. Konings, V. Rondinella, J. Somers, T. Wiss, *J. Nucl. Mater.* 319 (2003) 59–64.
- [42] A. Chartier, G. Catillon, J.-P. Crocombe, *Phys. Rev. Lett.* 102 (2009) 155503.
- [43] L. Fan, X. Shu, X. Lu, Y. Xie, Y. Ding, T. Duan, B. Dang, Y. Wu, *Ceram. Int.* 41 (2015) 11741–11747.
- [44] N. Laverov, S. Yudintsev, T. Livshits, S. Stefanovsky, A. Lukinykh, R. Ewing, *Geochim. Int.* 48 (2010) 1–14.
- [45] K. Yang, Y. Wang, P. Lei, T. Yao, D. Zhao, J. Lian, *J. Eur. Ceram. Soc.* 41 (2021) 6018–6028.
- [46] G. Pilania, K.R. Whittle, C. Jiang, R.W. Grimes, C.R. Stanek, K.E. Sickafus, B. P. Uberuaga, *Chem. Mater.* 29 (2017) 2574–2583.
- [47] B. Deng, Y. Zhang, *Chem. Phys.* 538 (2020) 110898.
- [48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [49] M. Rupp, A. Tkatchenko, K.-R. Müller, O.A. Von Lilienfeld, *Phys. Rev. Lett.* 108 (2012) 058301.
- [50] A. Stuke, M. Todorović, M. Rupp, C. Kunkel, K. Ghosh, L. Himanen, P. Rinke, *J. Chem. Phys.* 150 (2019) 204121.
- [51] A. Jain, G. Hautier, S.P. Ong, K. Persson, *J. Mater. Res.* 31 (2016) 977–994.
- [52] Y. Zhang, J. Duchi, M. Wainwright, *Conf. Learn. Theor.* (2013) 592–617.
- [53] B.J. Kennedy, B.A. Hunter, C. Howard, *J. Solid State Chem.* 130 (1997) 58–65.
- [54] E.J. Little Jr., M.M. Jones, *J. Chem. Educ.* 37 (1960) 231.
- [55] J.M. Farmer, L.A. Boatner, B.C. Chakoumakos, M.-H. Du, M.J. Lance, C.J. Rawl, J. C. Bryan, *J. Alloys Compounds* 605 (2014) 63–70.
- [56] R. Dronkowski, *Computational chemistry of solid state materials*, Wiley Online Library (2005).
- [57] B.C. Chakoumakos, *J. Solid State Chem.* 53 (1984) 120–129.
- [58] H. Yokoi, T. Matsui, H. Ohno, K. Kobayashi, *MRS Online Proc. Library* 353 (1994).
- [59] R. Ewing, W. Weber, F. Clinard Jr, *Prog. Nucl. Energy* 29 (1995) 63–127.
- [60] M.G. Brik, A.M. Srivastava, *J. Am. Ceram. Soc.* 95 (2012) 1454–1460.
- [61] C. Ma, J. Zhang, X. Chang, W. Hou, J. Wang, *Philos. Mag. Lett.* 88 (2008) 917–924.
- [62] J.P. Icenhower, W.J. Weber, N.J. Hess, S. Thevuthasan, B.D. Begg, B.P. McGrail, E. A. Rodriguez, J.L. Steele, *MRS Online Proc. Library* 757 (2002).