Article

Unveiling the effect of composition on nuclear waste immobilization glasses' durability by nonparametric machine learning

Check for updates

Yu Song[1], Xiaonan Lu [2], Kaixin Wang[1], Joseph V. Ryan[2], Morten M. Smedskjaer [3], John D. Vienna [2] ✉ & Mathieu Bauchy [1] ✉

Ensuring the long-term chemical durability of glasses is critical for nuclear waste immobilization operations. Durable glasses usually undergo qualification for disposal based on their response to standardized tests such as the product consistency test or the vapor hydration test (VHT). The VHT uses elevated temperature and water vapor to accelerate glass alteration and the formation of secondary phases. Understanding the relationship between glass composition and VHT response is of fundamental and practical interest. However, this relationship is complex, non-linear, and sometimes fairly variable, posing challenges in identifying the distinct effect of individual oxides on VHT response. Here, we leverage a dataset comprising 654 Hanford low-activity waste (LAW) glasses across a wide compositional envelope and employ various machine learning techniques to explore this relationship. We find that Gaussian process regression (GPR), a nonparametric regression method, yields the highest predictive accuracy. By utilizing the trained model, we discern the influence of each oxide on the glasses' VHT response. Moreover, we discuss the trade-off between underfitting and overfitting for extrapolating the material performance in the context of sparse and heterogeneous datasets.

The disposal of nuclear waste has far-reaching implications for both the environment and human well-being. To ensure safe storage, transportation, and disposal, vitrification of radioactive materials into waste forms is a crucial step in the waste management process[1,2]. Among the critical attributes for successful waste disposal, the chemical durability of waste-bearing glasses stands out as a paramount consideration, given the potential exposure of the disposal to aggressive environments[3]. Over the past decades, scientific research has demonstrated that the durability of waste glasses is fundamentally influenced, though not entirely determined, by their chemical compositions. Therefore, understanding the relationship between glass composition and durability performance holds special significance in the context of proper waste disposal.

At the Hanford Tank Waste Treatment and Immobilization Plant (WTP), the immobilized low-activity waste (ILAW) is slated for on-site disposal at the near-surface Integrated Disposal Facility[4]. As the ILAW glasses act as the primary barrier against the release of radionuclides to the biosphere, the Department of Energy (DOE) has mandated two chemical durability-related measurements to characterize the corrosion behavior of candidate ILAW glasses—the product consistency test (PCT)[5] and the Vapor Hydration Test (VHT)[6]. Different from the PCT which focuses on the steady-state leaching rate, the VHT is specifically designed to measure the glass corrosion rate under hydrothermal conditions, a process that leads to partial conversion of the glass into amorphous and crystalline alteration products. This hydrothermal corrosion test offers critical insights into the long-term durability of ILAW glasses in on-site disposal scenarios, contributing significantly to our understanding of the complex processes influencing glass performance in nuclear waste immobilization. The DOE has established specific requirements for measuring the VHT response limit as follows[7]:

The glass corrosion rate shall be measured using at least a seven (7)-day vapor hydration test run at 200 °C as defined in the DOE-concurred upon ILAW Product Compliance Plan. The measured glass alteration rate shall be

[1]Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), 5731B Boelter Hall, Department of Civil and Environmental Engineering, University of California Los Angeles, Los Angeles, CA 90095, USA. [2]Pacific Northwest National Laboratory, Richland, WA 99352, USA. [3]Department of Chemistry and Bioscience, Aalborg University, Aalborg 9220, Denmark. ✉e-mail: John.Vienna@pnnl.gov; bauchy@ucla.edu

1
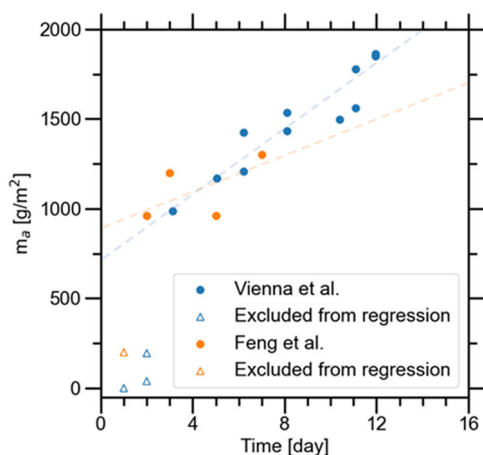
Fig. 1 | Impact of time on the amount of glass converted to alteration products in typical VHT tests[8,58,59]. The dashed lines represent linear fittings for the data measured from two studies, excluding the initial days.
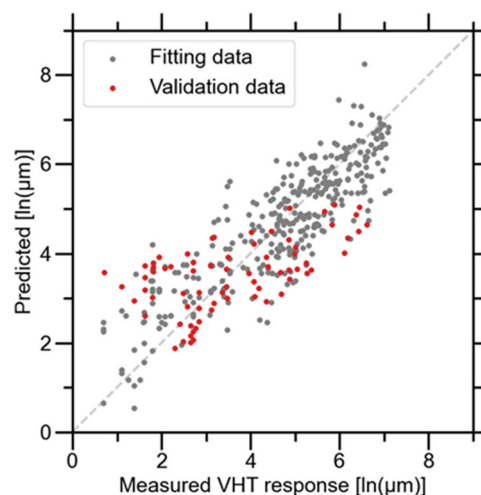
Fig. 2 | Predicted vs. measured logarithm VHT responses using PQM, aligned to an equivalent 24-day alteration depth[14]. For measured VHT response, the pooled standard deviation is 0.9034 ln(μm)[9], and, for predicted VHT response, the uncertainty (95% prediction interval) is illustrated in Supplementary Fig. 1.

less than 50 g/(m²·day). Qualification testing shall include glass samples subjected to representative waste form cooling curves. The vapor hydration test shall be conducted on waste form samples that are representative of the production glass.

In VHT, the rate of conversion is influenced by the coupled rates of glass corrosion and the formation of primarily crystalline zeolite alteration products. The nucleation of zeolite phases in this process is sporadic, leading to a high degree of response variability, particularly during the early stages of conversion (Fig. 1)[8]. Within the typical 24-day VHT duration, glasses with high alteration rates have usually progressed past the early sporadic phase, while glasses with low alteration rates may still be within the early sporadic phase. Consequently, simple functions used to describe the VHT responses of waste glasses are inaccurate. The high degree of nonlinearity and the variability in zeolite nucleation, coupled with variability in experimental factors, contribute to high bias and uncertainty in VHT response prediction. Jiricka et al. highlighted the significant impacts of surface finish, volume of water, and cutting procedures on the single-time VHT responses, all of which are generally not strictly controlled across different VHT tests[8]. As a result of these experimental variations, a relative standard deviation of 63% was observed for the pooled replicated single-time VHT response[9].

In real production scenarios, glass property-composition models are utilized to predict the VHT response of glasses during the processing of each waste batch at Hanford's LAW vitrification facility. The use of these models, along with their uncertainty descriptions, is necessary to ensure with sufficient confidence that glasses satisfy the VHT specifications listed in the WTP contract[10]. This approach is also used for other glass processing and product quality-related properties[11]. Existing VHT response models have met this goal with limited success. In terms of the previous modeling efforts, Vienna et al. modeled the logarithm VHT response of 58 glasses as a first-order function of composition[12]:

$$\ln\left(r, \frac{\text{g}}{\text{m}^2 \cdot \text{d}}\right) = \sum_{i=1}^{q} r_i g_i \tag{1}$$

where r is the VHT rate at 200 °C (based on multi-time VHT responses), r_i is the ith component coefficient, and g_i is the ith component mass fraction in a glass. More recently, partial quadratic mixture (PQM) models were applied to advance the prediction of the VHT response in a similar manner[13]:

$$\ln\left(d, \mu\text{m}\right) = \sum_{i=1}^{q} v_i g_i + \text{Selected}\left\{\sum_{i=1}^{q} v_{ii} g_i^2 + \sum_{i=1}^{q-1} \sum_{j=i+1}^{q} v_{ij} g_i g_j \dots\right\} \tag{2}$$

where d is the VHT alteration thickness in μm after a 24-day test duration at 200 °C, v_i is the ith component coefficient, v_{ii} is the ith component squared coefficient and v_{ij} is the ith and jth component cross-product coefficient. These models demonstrated some success in predicting VHT responses within relatively narrow composition regions but faced high relative uncertainty. For instance, to satisfy the WTP constraint of the VHT rate being 50 g/m²/d with 90% confidence, the predicted rate needed to be below 7.9 g/m²/d[10]. Also, the validation performance of the model is unsatisfactory, as shown in Supplementary Fig. 1. As the composition regions expanded, particularly for the relatively large region of enhanced LAW glasses[14–16], the prediction uncertainty increased, and distinct biases emerged. The PQM model by Vienna et al., for instance, over-predicted low VHT responses while under-predicting higher VHT responses (Fig. 2). Vienna et al. also attempted to address this issue using an artificial neural network (ANN) to model VHT response as a function of composition[15]. While this model exhibited reduced uncertainty, it did not perform well on validation data, indicating overfitting. Consequently, Vienna et al. did not obtain a suitable model to quantify VHT responses in enhanced ILAW glass data and so developed a logistic regression model to predict pass/fail state of glasses relative to the 50 g m⁻² d⁻¹ contract limit[16]. Therefore, for both research and the design of LAW glasses with enhanced performance, a more robust model is needed to overcome experiential noise and capture the true compositional effects on VHT response.

This study leverages advanced machine learning methods to unveil the distinct effect of different oxides on the durability performance of nuclear waste immobilization glasses. We utilize a dataset comprising VHT alteration rates of more than 600 Hanford LAW glasses, encompassing a wide compositional envelope. Four machine learning methods, including PQM and ANN from previous studies, and two nonparametric approaches, namely, local linear regression (LLR) and Gaussian process regression (GPR), are employed to predict their VHT responses. The nonparametric methods, being more capable of fitting complex relationships without presuming the target function's form, typically offer superior performance[17]. Our test results demonstrate that GPR achieves the best overall accuracy and robustness in predicting the VHT response of unseen glass compositions, significantly outperforming the benchmark models. Through detailed interrogation of the GPR model, we decompose the complex relationship between glass composition and VHT response, effectively unveiling the individual effects of the different major oxides. Based on the observations from this study, we further discuss an interesting phenomenon that is generic to machine-learning-based material studies, where a slightly-
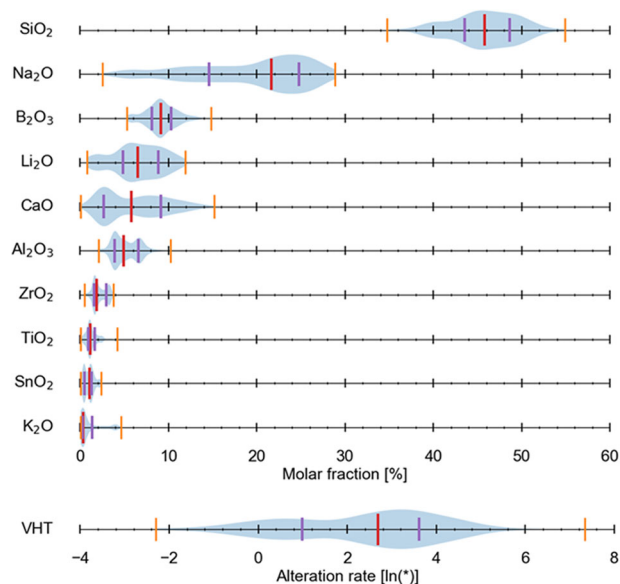
Fig. 3 | Composition ranges of the ten oxides in the 585 considered nuclear waste immobilization glasses for modeling the VHT responses, along with the distribution of the VHT responses in the dataset. Note that the unit of the scaled VHT alteration rate, *, represents g/(m² d). The red, two purple, and two orange whiskers indicate the median, lower and upper quartiles, and lower and upper extremes of the feature values, respectively.

Table 1 | Comparison of the predictive accuracies and lack-of-fit statistics for four models—PQM, ANN, LLR, and GPR

		PQM	ANN	LLR	GPR
Training set	R^2	0.62 ± 0.01	0.78 ± 0.01	0.77 ± 0.18	0.76 ± 0.01
	RMSE [ln(*)]	1.05 ± j0.01	0.80 ± 0.02	0.66 ± 0.16	0.83 ± 0.02
Test set	R^2	0.58 ± 0.04	0.63 ± 0.08	0.61 ± 0.15	0.65 ± 0.07
	RMSE [ln(*)]	1.09 ± 0.05	1.03 ± 0.12	0.93 ± 0.23	1.00 ± 0.12
Lack of fit, F		1.50 ± 0.00	0.87 ± 0.03	1.28 ± 0.01	0.63 ± 0.04

The mean and standard deviation values are calculated based on the model performance over ten train-test splits. Here, * represents the unit of the scaled VHT alteration rate, i.e., g/(m²·d). Lack of fit statistics are calculated based on all samples in the dataset.

overfitted model is found to be beneficial for extrapolating a small dataset with a heterogeneous sample distribution.

Results

Prediction accuracy of the machine learning models

Based on the data reported by PNNL-30932 report[16], a dataset consisting of 639 single-time VHT responses was curated. This dataset considers the influence of ten major oxides (as the model inputs) on the VHT response (as the model output). The distributions of those inputs and output are summarized in Fig. 3. As indicated by the orange whiskers in Fig. 3, the glass compositions involved in this dataset range over a broad envelope as relevant to diverse glass waste forms. Additional details of the data curation are provided in the "Methods" section.

Based on the procedures described in the "Methods" section, the four models—PQM, ANN, LLR, and GPR—were optimized. The optimal prediction accuracies of these models were compared using ten different train-test splits, and the results are presented in Table 1. Additionally, Fig. 4 showcases the predicted and true VHT alteration rates of the test samples from one single train-test split. Overall, all the models achieve decent prediction accuracy in mapping the glass composition to the VHT alteration rate. While the averaged RMSE values on the test set ~1.0 ln[g/(m²·d)] for all four models, it is worth noting that the pooled standard deviation (SD) is found to be 0.9034 ln[g/(m²·d)]) from analyzing duplicate glasses (i.e., same glass composition with multiple measured values), representing the uncertainties caused by fabricating glasses, and performing VHT analyses[9]. The small difference between RMSE and pooled SD indicates that the models are properly fitted (i.e., not over or underfitting).

Another noteworthy observation from Fig. 4 is that all four models exhibit decreased prediction accuracy as the VHT alteration rates decrease. This trend can be attributed to the reduced number of samples with very low VHT alteration rates in the dataset, as illustrated in Fig. 3. Such scarcity of samples in this range implies that corresponding predictions entail more extrapolation. In that regard, an intriguing phenomenon observed during model optimization is the delicate balance between the degree of overfitting and extrapolability, influenced by varying degrees of model regularization. This dynamic arises due to the uneven distribution of samples in the dataset.

Specifically, a slightly-overfitted model extrapolates more accurately for glass compositions with very high or very low VHT alteration rates. This point is discussed in detail in the "Discussion" section.

Among the four models, the PQM model, despite achieving a higher accuracy than reported in the previous study[14], exhibits the lowest performance. The relatively lower accuracy of the PQM model can be attributed to its finite analytical terms, which struggle to adequately capture the intricate dependence between the glass composition and its VHT response[18]. For further insights into the analytical terms of this updated PQM model, please refer to Supplementary Note 3. Conversely, the ANN model demonstrates considerable improvements in both test R^2 and RMSE, owing to its higher flexibility in fitting non-linear patterns. However, a significant gap is observed between the training and test accuracies of the ANN model, suggesting a tendency to overfit. This observation is consistent with previous findings and raises concerns about the ANN model's ability to generalize well on new samples[14], particularly for new glasses located near the edge of the compositional envelope.

In comparison, both the nonparametric models exhibit superior prediction accuracies. The LLR model shows the best performance on test RMSE among the four models, but the high prediction variations of LLR (indicated by the standard deviation values in Table 1) raise questions about its robustness. The high variations of LLR can be attributed to the fact that some samples in the VHT dataset are sparsely distributed in the compositional envelope, causing the local fitting of the LLR model to vary significantly across the different train-test splits. In contrast, the GPR model achieves a good prediction accuracy, with the highest R^2 accuracy and the second-lowest RMSE on the test set. Importantly, the GPR model does not involve the instability associated with LLR or the overfitting issue of ANN. Furthermore, the lack of fit test indicates that GPR has the smallest F score (i.e., least likely to involve the lack-of-fit issue) compared to PQM and LLR models. Considering all the above-discussed factors, GPR emerges as the most suitable modeling method for the logarithm VHT response dataset[19]. Consequently, results and discussion in the subsequent sections focus only on GPR.

One-dimensional effect of individual oxides on VHT response

To interpret the effect of individual oxides on the VHT alteration rate as learned by the optimized GPR model, we conducted a one-dimensional feature effect analysis. We first assumed a hypothetical glass composition based on the median value of each oxide feature (as shown in Fig. 3), representing a generic composition of nuclear waste immobilization glasses. Subsequently, we analyzed the effect of each oxide by sequentially jittering its content within its 15-to-85 percentiles in the dataset, while rescaling other oxides such that the total oxide content remained unchanged. The impact of each oxide was then evaluated based on the variation in the predicted VHT alteration rate resulting from the jittered oxide content.

The results of the individual oxide effect analysis are summarized in Fig. 5. Notably, the different oxides exhibit distinct effects on the VHT

Fig. 4 | Comparison of the accuracy of the machine learning models. Predicted vs. measured VHT alteration rate for (**a**) PQM, (**b**) ANN, (**c**) LLR, and (**d**) GPR models, based on the same test samples from a single train-test split. The points are color-coded based on the density of the scatters on the plot (the concentration increases from blue to red), and * represents the unit of the scaled VHT alteration rate, i.e., $g/(m^2 \cdot d)$. The dashed line is added to visualize the perfect agreement.
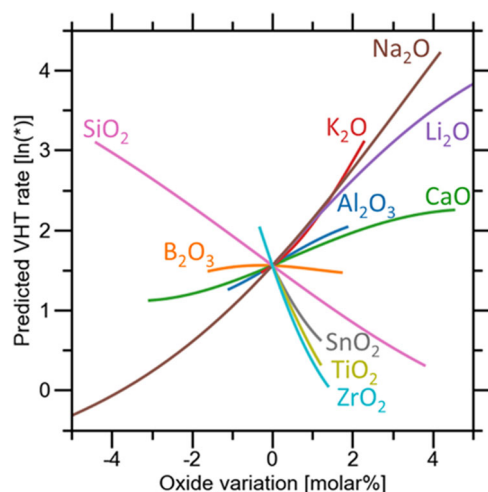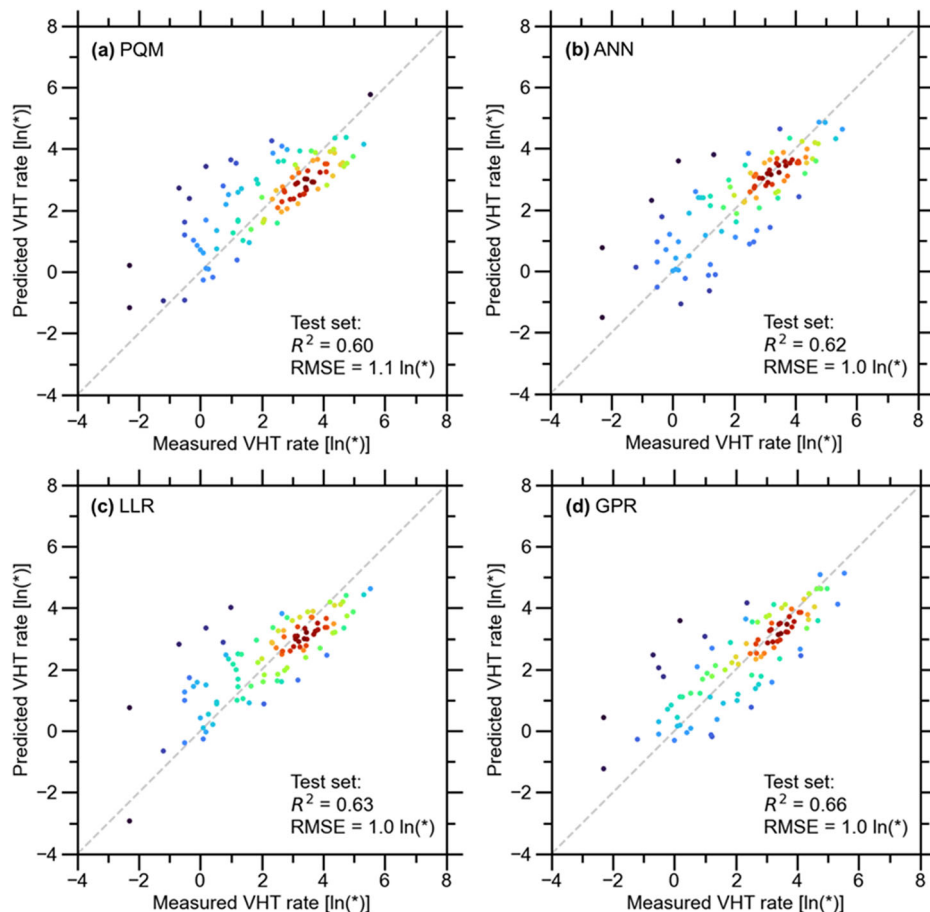
Fig. 5 | One-dimensional composition effect on VHT alteration rate based on the optimized GPR model. The centroid corresponds to the median value of each oxide feature in the dataset. Note that* represents the unit of the scaled VHT alteration rate, i.e., $g/(m^2 \cdot d)$.

response, characterized by varying slopes and degrees of nonlinearity, which broadly corroborate the observations reported in a prior analysis utilizing the PQM model[16]. Based on the slopes of the feature lines, the ten oxides considered in this study can be categorized into two groups (ranked by the feature slope and the quantity of each oxide in the dataset)—those that increase the VHT alteration rate of the glass (Na_2O, Li_2O, K_2O, CaO, and

Al_2O_3), and those that decrease the rate (ZrO_2, SiO_2, TiO_2, and SnO_2). Note that B_2O_3 has minimal impact on VHT response. Meanwhile, we stress that the displayed feature effects can be influenced by the choice of the reference composition (i.e., the centroid in Fig. 5). The effect of the same oxide can also vary depending on the reference composition, which, in turn, highlights the challenge of interpreting the effect of each individual oxide with a conventional model.

Model interrogation with SHAP analysis

To provide a more comprehensive understanding of the prediction of VHT alteration rate by the optimized GPR model, we employed the SHAP analysis, a powerful model interpretation technique[20,21], as briefly reviewed in the "Methods" section. Unlike the one-dimensional composition analysis, the SHAP interpretation is based on the actual compositions in the VHT dataset, without relying on assumptions about a reference glass composition. For this analysis, all the samples in the dataset were considered, ensuring a robust and detailed examination of the model's predictions.

In Fig. 6, we present a summary of the marginal contribution of each oxide component to the VHT alteration rate. The oxides are here ranked based on their absolute influence on the VHT alteration rate as captured by the Shapley values. The horizontal axis represents the Shapley value (i.e., relative impact), where positive (negative) values indicate an increase (decrease) in the VHT rate. For each oxide, the horizontal distribution of the scatters shows the variation of its marginal contribution in response to the standardized feature values such that all oxides can be compared with their mean value aligned to the centerline in the plot. This enables easy comparison of all oxides with their mean values. Taking Na_2O as an example, this plot suggests that, as the content of this oxide increases (as the scatter color shifts from blue to pink), its marginal contribution changes from negative to
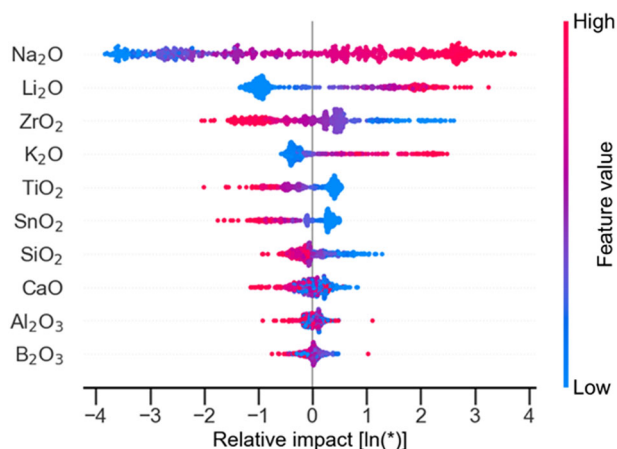
Fig. 6 | Summary of the SHAP analysis. This plot shows the relative impact (i.e., marginal contribution) of the oxide components on the VHT alteration rate, as ranked by importance from top to bottom. Note that the feature values in this plot have been standardized to ensure a fair comparison. * represents the unit of the scaled VHT alteration rate, i.e., $g/(m^2 \cdot d)$.

positive. This suggests that a higher amount of Na_2O tends to raise the VHT alteration rate while low concentrations show the opposite effect.

We next display the feature-wise SHAP results in Fig. 7, grouping the oxides based on the slopes of their feature effects from Fig. 5. In each plot, the actual content of each oxide is plotted against the marginal contribution, with each point representing a specific sample in the VHT dataset. We observe that most of the oxides exhibit a monotonous influence on the VHT rate. Among all, Na_2O, Li_2O, ZrO_2, and K_2O, show clear effects in affecting the VHT rate, with ZrO_2 showing a negative influence opposite to the other three oxides. In contrast, the data patterns extracted from CaO and Al_2O_3 are less clear in both Figs. 6 and 7. This point is further discussed with additional evidence based on the SHAP analysis in the "Discussion" section.

Discussion

We first discuss the trade-off between model overfitting and extrapolability associated with the uneven sample distribution in the dataset, as noted in the "Results" section. This is a common challenge faced in many machine-learning-based material studies due to the limited scale of common lab testing[22–25]. Conventionally, it is expected that a model's extrapolation capability is maximized when overfitting is minimized. However, our findings based on the VHT dataset reveal a somewhat surprising outcome: a slightly-overfitted model appears to exhibit more accurate extrapolation at very low or very high VHT rates.

This trade-off issue is collectively illustrated with Figs. 8, 9, and 10, as explained in the following. We first employ progressive regularization of the GPR model, transitioning from overfitting to underfitting, by gradually increasing the noise level parameter. Using the RMSE difference between the training and test set samples as a measure, we then track the model deviation at different ranges of the VHT alteration rate. As a comparison, Fig. 8 presents the prediction deviations between a non-overfitted model (achieved by minimizing the difference between the train and test R^2 accuracies) and a slightly-overfitted model. For statistical representativeness, the single sample with a VHT rate larger than 6 $\ln[g/(m^2 \cdot d)]$ was excluded from this analysis.

Ideally, a model with robust extrapolability should exhibit a consistent level of deviation across the entire range of the VHT rate. In Fig. 8a, the non-overfitted model demonstrates relatively low deviations within the interpolation region, but its deviation significantly increases in the extrapolation regions. Conversely, the slightly-overfitted model in Fig. 8b shows marginally inferior interpolation performance, yet it achieves substantially lower deviations during extrapolation. Consequently, compared to the non-
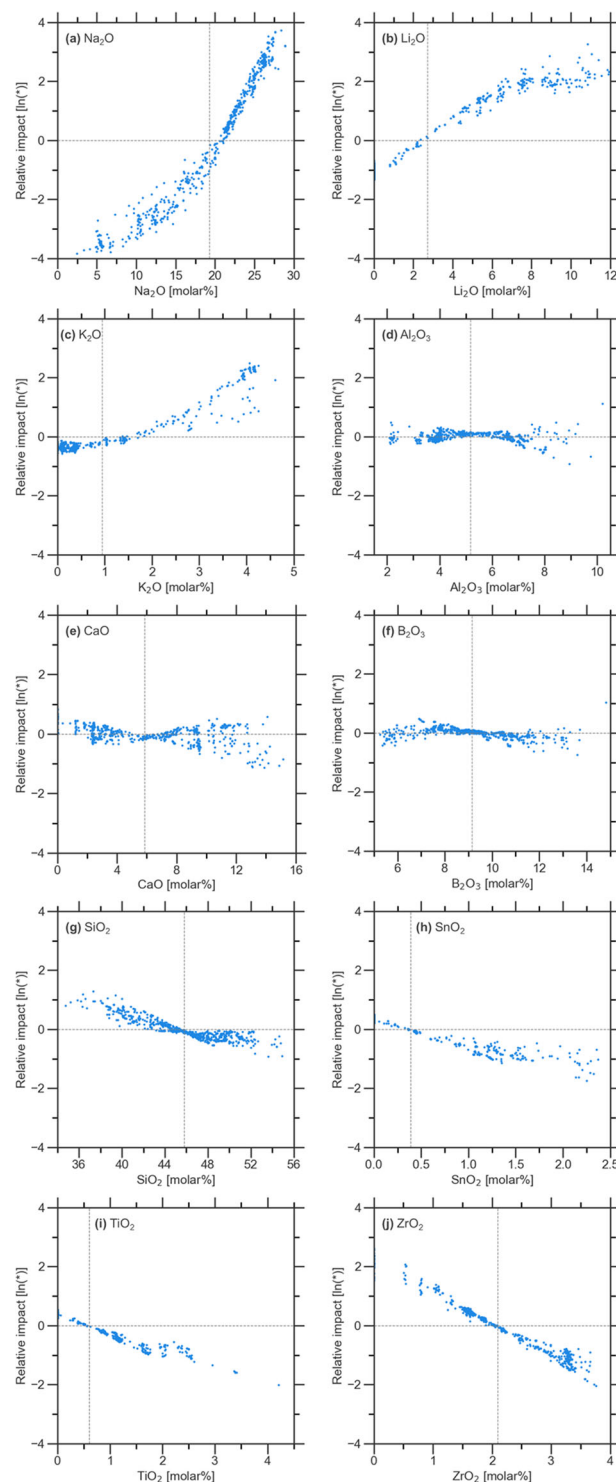

Fig. 7 | Feature-wise results of the SHAP analysis. Those plots show the correlation between the oxide content (on the actual scale) and marginal contribution: **a** Na_2O, **b** Li_2O, **c** K_2O, **d** Al_2O_3, **e** CaO, **f** B_2O_3, **g** SiO_2, **h** SnO_2, **i** TiO_2, and **j** ZrO_2. Here, the feature values are based on the actual scale. To assist comparison, the vertical and horizontal dash lines indicate the mean feature value and zero marginal contribution, respectively.

overfitted model, the slightly-overfitted model appears to be better suited for predicting glass compositions with extreme VHT responses.

To further investigate the trade-off issue illustrated in Fig. 8, the average deviation (mean deviation averaged from the different VHT intervals), and the degree of overfitting (i.e., the RMSE difference between the training and

Fig. 8 | Trade-off between model overfitting and extrapolability observed in this study. Variation of the prediction deviation (i.e., the RMSE difference between the training and test sets) at different VHT alteration rates, based on the (**a**) non-overfitted and (**b**) slight-overfitted GPR models. The interpolation and extrapolation ranges are roughly divided according to the data distribution in the VHT dataset. Note that * represents the unit of the scaled VHT alteration rate, i.e., g/(m^2·d).
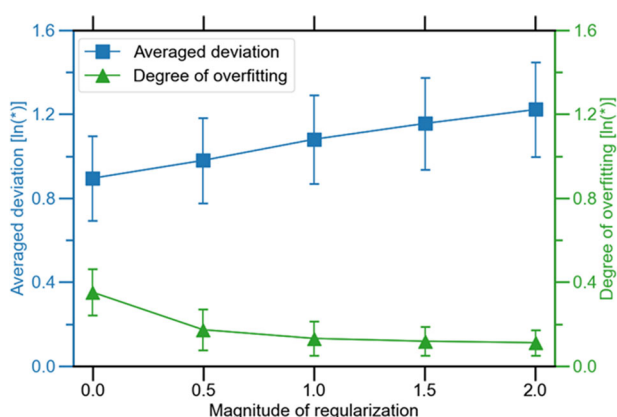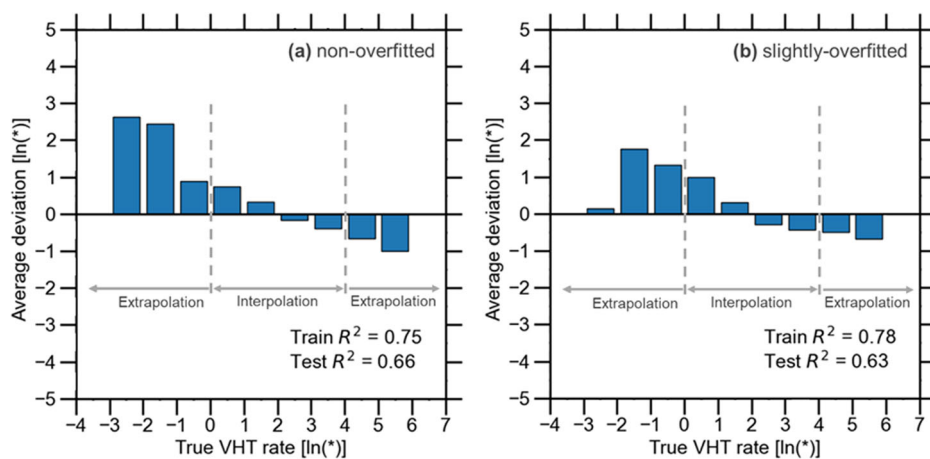
Fig. 9 | Variations of the prediction deviation (blue squares) and the degree of overfitting under different levels of model regularization (green triangles). Here, the two metrics are both calculated based on the RMSE difference between the training and test set samples, but the former takes the average from the different VHT intervals (see Fig. 8) while the latter is on the overall difference.

test set samples) at different levels of model regularization are plotted in Fig. 9. Noted that the zero regularization here indicates a very low level of regularization that causes the model to overfit, as the noise level is not the only parameter controlling the regularization in GPR. Our analysis reveals that, as the model regularization increases, the degree of overfitting gradually diminishes as anticipated, but this is accompanied by a continuous increase of the average deviation, which indicates a decline in the model extrapolability. As seen in Fig. 8, the rise of the average deviation is primarily driven by the higher deviations at low and high VHT rates. These trends, depicted in Figs. 8 and 9, are not commonly expected, as the general understanding is that overfitting leads to larger deviations, undermining model performance.

The superior performance of the slightly-overfitted model observed in this study is likely related to the uneven distribution of the VHT alteration rate in the adopted dataset. Specifically, most of the VHT samples are concentrated within the mid-range of the alteration rate between 0 and 4 ln[g/(m^2·d)] (see Fig. 3), while fewer samples are outside this range. The observed issue is illustrated with a simplified case of the heterogeneous dataset in Fig. 10. In this figure, the majority of the training samples are concentrated at 4 to 8, and the major prediction deviations from different models are indicated with arrows. The model's extrapolability is reflected by the magnitude of deviations at the two ends, where both the non-overfitted and overfitted models

exhibit large deviations when predicting samples outside the middle range. In contrast, despite the slightly higher variance compared to the non-overfitted model, the slightly-overfitted model achieves more reasonable predictions when extrapolating beyond the populated region.

Returning to the VHT dataset, it is apparent that the model optimization is dominated by the model accuracy in predicting the most populated samples within the mid-range of VHT rate. Consequently, much lower weights are assigned to extrapolating glass compositions with high and low VHT rates. With slightly lower regularization, however, more weight of the model optimization is allocated to the less populated regions. As a result, the slightly-overfitted model extrapolates the VHT rate substantially better at the extremities.

Drawing from the observations made from the VHT dataset as discussed above, different strategies can be employed to select the best model based on the target of the model prediction. If the model is primarily intended for interpolating material performance within the populated region of the dataset, the non-overfitted model is expected to provide the best predictability. However, in cases where there is an interest in extrapolating material performance, adopting a slightly-overfitted model could yield more robust predictions for extrapolating material behavior. Indeed, the ability of the slightly-overfitted model to accurately extrapolate at very low or very high VHT rates is particularly relevant and significant for material research. This is because the slightly-overfitted model may result in more robust predictions in extrapolating the material behavior, which facilitates the exploration of unexplored regions in the compositional space, potentially leading to the discovery of new materials with unconventional compositions.

Next, we discuss the effect of individual oxides on the VHT alteration rate, as conducted in both the one-dimensional feature effect analysis and the SHAP analysis, yields consistent and insightful results. Based on the results in Figs. 5, 6, and 7, the influence of the ten oxides can be ranked descendingly as follows: Na_2O, Li_2O, ZrO_2, K_2O, TiO_2, SnO_2, SiO_2, CaO, Al_2O_3, and B_2O_3. Among all, Na_2O has the most substantial impact on the VHT alteration rate, with higher concentrations leading to a faster rate of alteration, which becomes more pronounced when its molar fraction exceeds 20%[14,16]. Meanwhile, only CaO and Al_2O_3 exhibit unclear effects on affecting the VHT rate. In general, the interpretation of these oxide effects aligns well with the roles of the corresponding atoms in altering the glass's formation ability and chemical durability, which are related through the oxide composition and atomic structure. In terms of the VHT test, the response of a glass is determined by two primary factors: 1) the influence of the components on glass network stability and 2) the influence of the components on the stability of the solution leading to the driving force to dissolution.

Fig. 10 | Illustration of the trade-off between overfit and model extrapolability. Predictions made by (**a**) non-overfit and slight-overfit and (**b**) strong-overfit models trained based a simplified heterogeneous dataset. In both cases, the arrows indicate the major prediction deviations from the true function.
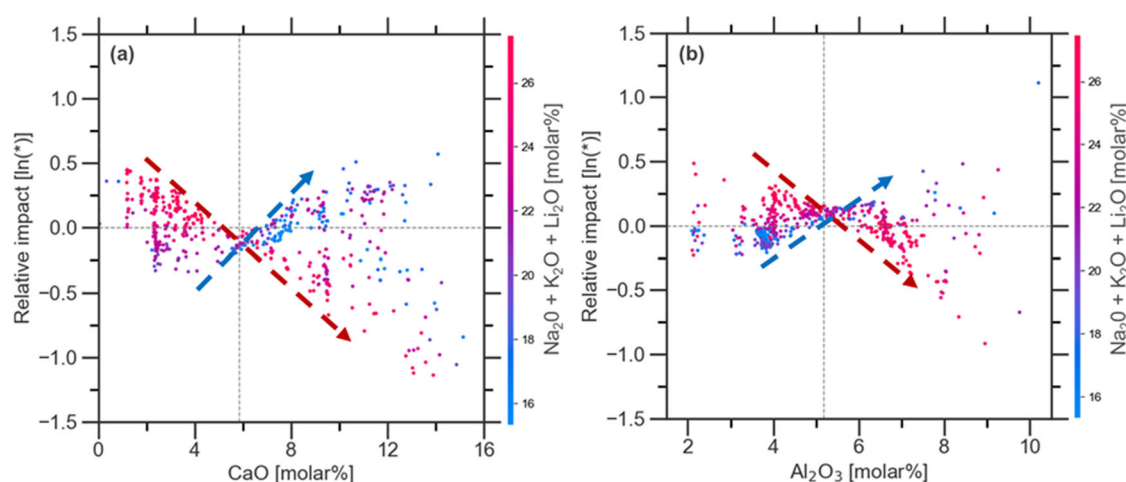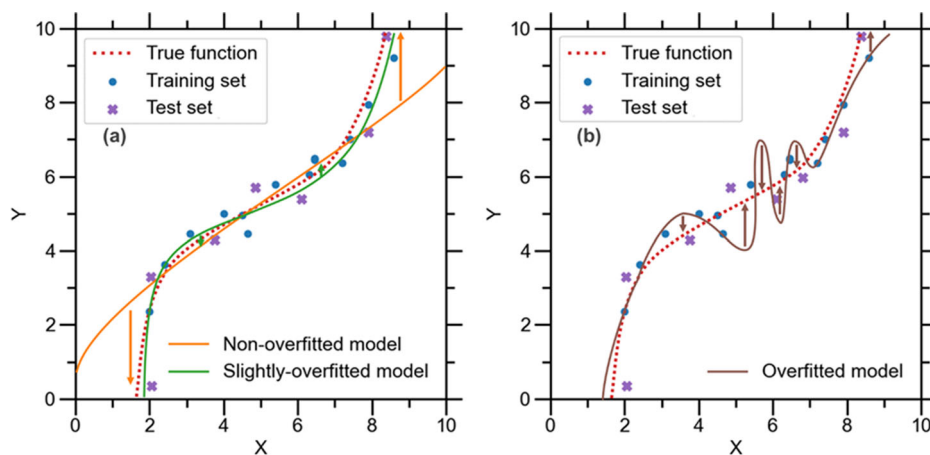
Fig. 11 | Interactive effects between oxides as revealed by the SHAP analysis. Variations on the marginal contribution of (**a**) CaO and (**b**) Al_2O_3 to VHT response under the influence of the total content of alkaline oxides ($Na_2O + K_2O + Li_2O$). Here, the arrows are added to highlight the dual effects of these two oxides, and the feature values are based on the actual scale. To assist comparison, the vertical and horizontal dash lines indicate the mean feature value and zero marginal contribution, respectively.

Regarding the glass network stability, it is associated with the role of different atoms in altering the topology of the glass atomic network (see topological constraint theory)[26–28], which influences the strength of the chemical bonds within the glass[29,30]. Here, Na, Li, and K atoms that form alkali oxides play a role in depolymerizing the glass network, thus reducing the corrosion resistance of the nuclear waste immobilization glass and therefore increasing the VHT rate. In contrast, Si and B atoms, as typical network-forming elements, enhance the glass formation ability and, consequently, reduce the VHT response. Additionally, the weaker network-modifying elements, Zr, Ti, and Sn, also contribute to enhancing glass network formation[31]. Aluminum has a mixed effect on the glass structure depending on the overall glass bonding environment[30,32].

As for the effect of the solution on glass dissolution, alkali ions released into the solution increase its pH, driving a higher dissolution rate while boron buffers the pH of the corroding solution. Zirconium, Ti, and Sn are only sparingly soluble in solution and therefore have little effect on this mechanism. Orthosilicic acid and aluminate ion concentrations in solutions most strongly influence the driving force for dissolution[33,34]. High Al concentrations in a caustic solution cause a precipitation of zeolites which dramatically reduce the concentrations of both aqueous Si and Al, then increase the dissolution rate[35]. Notably, the effects of Ca and Al atoms on the VHT response appear to be mixed, with their marginal contributions being either positive or negative, even at fixed oxide content levels (see Fig. 7e, d).

By further taking advantage of the SHAP analysis, we revisit the effect of those oxides while focusing on identifying their most interactive features[21], as displayed in Fig. 11. We find that the total content of alkaline oxides (Na_2O, K_2O, and Li_2O) has the strongest interactive effect with CaO and Al_2O_3. Under low alkaline oxide content, the marginal contributions of CaO and Al_2O_3 are positive, while this trend reverses with the presence of a high total alkaline content. This dual pattern may be linked to the significant impact of alkaline oxides on solution pH, which is known to accelerate glass corrosion[32]. This observation points to a potential feedback effect involving the solution chemistry that influences the VHT response in complex ways. Thus, CaO and Al_2O_3 warrant further investigation to understand the underlying solution feedback mechanisms that govern their influence on the VHT response. To systematically analyze the interactive effects among other oxides that also exhibit potential significance in glass alteration processes, future investigations may benefit from employing more sophisticated criteria and validating the results through experimental characterizations.

Overall, machine learning proves to be a valuable tool for understanding the relationship between the composition of nuclear waste immobilization glasses and their VHT responses. Through proper model interpretations, data-driven machine learning offers meaningful insights into the durability of these glasses and the effects of individual oxides, which are not easily accessible using conventional approaches. The findings presented herein contribute to the knowledge and development of nuclear

Table 2 | Outlying data removed from the VHT model fitting dataset[16]

Reason for removal	Glass IDs of the removed samples
Noise in composition (mass fraction of minor oxides >1.9%)	LAWA46, LAWA47, LAWA48, LAWA64, LAWABP1, LAWA54, LAWA55, LAWA58, LAWA59
High MgO content (mass fraction >6.0%)	ORPLA28, ORPLA29, ORPLA31, ORPLA32
Fitting outlier	New-OL-57284
VHT rate 7d > 24d	LP2-IL-04

waste immobilization glasses, paving the way for enhanced material design and performance optimization in this critical field of research. More importantly, the generic analyses presented in this study can serve as a paradigm for other machine-learning-based material investigations.

Methods

Curation of the VHT dataset

A comprehensive database consisting of 654 single-time VHT responses, measured for 595 individual glasses, was initially compiled, including 59 replicates of 44 distinct glass compositions[8,14,15,36]. The raw data is documented in the Appendix A of the PNNL-30932 report[16]. The majority of the VHT results were obtained at the standard 24-day measurement period. In cases where some glasses were completely corroded during the 24-day duration, shorter measurement times such as a 7-day test were employed. For the sake of uniform analysis, those shorter testing points were extrapolated to an equivalent 24-day specimen thickness, assuming a constant rate of alteration from the test's initiation. It is worth noting that this approach, while ensuring consistency, may introduce some uncertainty to the measured value.

Following an initial screening process, a total of 15 glasses were excluded from the database, as summarized in Table 2. The removal criteria were as follows (i) extreme compositions for 13 glasses, (ii) consistently gross model fit outliers for one glass, and (iii) abnormal VHT measurement for one glass. For the first type of outliers, the removal of the extreme compositions was based on a systematic analysis previously reported by Vienna et al. (see Section 4.1.1.1 of the cited report)[16]. For the second type, a single glass consistently appeared as an isolated data point, residing far from all other data points in the fitting conducted by different models. Regarding the third type, the 7-day corrosion result exceeded the 24-day result, possibly indicating an experimental issue such as water loss during the test or reflux.

After removing the outliers, a dataset was prepared for the subsequent machine learning analysis comprising 639 VHT responses from 585 individual glass compositions, with an additional 54 replicates of 42 glass compositions. Given the limited size of this dataset, some oxides in the glass compositions that are less influential to the VHT alternation rate were omitted to reduce data dimensionality, thereby improving the efficiency of the machine learning model training and mitigating the "curse of dimensionality" phenomenon[37]. The identification of uninfluential oxides was previously documented by Vienna et al. in a comprehensive study (refer to Section 3.3 of the cited report)[14]. This analysis initially pinpointed 15 oxides with reasonably distributed concentrations for potential inclusion as model inputs. These included Al_2O_3, B_2O_3, CaO, Fe_2O_3, K_2O, Li_2O, MgO, Na_2O, P_2O_5, SiO_2, SnO_2, TiO_2, V_2O_5, ZnO, and ZrO_2. Through a systematic elimination process based on the step-by-step removal of oxides with negligible effects, four oxides— MgO, P_2O_5, V_2O_5, and ZnO—were excluded from further consideration. Herein, Fe_2O_3 was further excluded due to its strong correlation with multiple oxides known to exert a significant influence on the VHT alteration rate, such as Al_2O_3, Na_2O, SiO_2, and ZrO_2. Consequently, a total of ten major oxide components (see Fig. 3) were selected based on their impact on VHT response[14]. These components constituted a significant portion of the total molar content, ranging from 82.2% to 96.5% with a median of 92.5%. For the model inputs, the molar fractions of the selected oxide were further normalized to sum to one.

Regarding the model's output, the raw VHT response of the selected samples ranged from 0.7 to 13848 μm in depth. To standardize the influence of different sample sizes and testing periods, the VHT response was scaled as the glass mass change per square meter per day. Subsequently, the model's output was taken as the natural logarithm of the VHT alteration rate, reducing the skewness of the output distribution. Following these transformations, the VHT alteration rate now ranges from −2.30 to 7.33 ln[g/(m^2·d)] in the dataset. Note that the ranges of both the glass composition and VHT alteration rate considered herein are broader than those in previous investigations[8,14,15,36], which presents additional challenges in modeling the relationship between glass composition and VHT alteration rate.

Machine learning analyses

Based on the dataset, we employed various machine learning approaches to predict the VHT alteration rate based on the glass composition as input. In detail, we considered four regression methods, namely, PQM, ANN, LLR, and GPR. The first two have been used in previous studies and they are considered as baselines[15,18], while the latter two are nonparametric models that may exhibit superior performance in fitting the waste glass dataset. To ensure a fair comparison, all four models were provided with the same ten input features for model training (see Fig. 3). The modeling work was conducted in Spyder IDE (5.4.3) with Python (3.10.10), and the specific program packages (if applicable) used are provided below. A brief overview of these four models is provided as follows.

The PQM model employs a combination of relatively simple terms, including linear, squared, and quadratic cross-products, to generate predictions[38]. By incorporating squared and quadratic terms, PQM can capture certain non-linear patterns in the data. Previous studies have demonstrated its feasibility for modeling the VHT response of nuclear waste immobilization glasses[13]. The PQM model in this study was built using the statsmodels package[39].

The ANN model has been widely popular in machine learning due to its flexibility in processing diverse datasets and performing various machine learning tasks[40–42]. In this study, the ANN model was built using the PyTorch package[43]. To optimize the model's architecture, considering the limited size of the VHT dataset and experience established from previous studies[22,44,45], we established a network with eight and four artificial neurons in two hidden layers, respectively. ReLU activation and Batch Normalization were implemented between each layer to enhance model accuracy[46]. Furthermore, Adam optimizer and mean square error (MSE) loss were adopted for guiding the model optimization.

The LLR model performs linear regression locally in the data space, providing non-linear and smooth predictions on a global scale[47]. LLR employs a distance-based weight function to determine the influence of each data point on local fitting, making it nonparametric and highly expressive for modeling complex relationships. LLR also exhibits good resistance to noise in the dataset. The LLR model in this study was directly coded according to the algorithm.

The GPR model is a probabilistic-based prediction method established on Bayes' Rule[48]. GPR assumes that both the input features and output target follow Gaussian processes, enabling the modeling of the probabilistic distribution of the output target (i.e., *posterior* distribution) based on the input features (i.e., *prior* distribution)[49]. This allows GPR to provide not only predictions (mean of the probabilistic distribution) but also estimates of prediction confidence (variation of the probabilistic distribution). Being nonparametric, GPR can fit completely unknown relationships without assuming a specific mapping function. GPR requires selecting an optimal

kernel function, which should ideally echo the generic trend of the predicted output distribution. The GPR model in this study was built using the GpyTorch package[43,50]. Based on the kernel selection pipeline established in a previous GPR study[17], we used a combination of Linear and Radial Basis Function (RBF) kernels for modeling the VHT dataset. The former ensures good model extrapolability, while the latter allows the model to fit complex data patterns[51]. Another hyperparameter, noise level, was tuned for model regularization to prevent overfitting. Further details on the noise level choice are provided in the "Results" section.

In terms of the optimization of the four models, we primarily employed brute-force searches to identify their respective optimal hyperparameters, based on ten independent random training repetitions to evaluate model performance. Further details about the model tunning and the selected optimal hyperparameters are provided in Supplementary Note 2. The search was performed on 85% of the samples (i.e., training set), which were randomly stratified from the dataset to ensure a statistically consistent distribution within the raw dataset[22,44,52]. The remaining 15% samples (i.e., test set) were held hidden in order to test the true performance of the models with optimized hyperparameters since those samples were never exposed to model training. In each repetition, 15% of the training set samples were held out as a validation set to determine the optimal hyperparameters (i.e., those yielding the lowest averaged MSE loss). Model performance comparison was based on two statistical metrics computed on the test samples: coefficient of determination, R^2, and root mean square error, RMSE[46]. Additionally, a lack-of-fit F-test was conducted to assess whether prediction errors were due to lack of model fit[53], where a high F score (>1) indicates a lack-of-fit model.

SHAP analysis

Advanced machine learning algorithms like ANN can effectively interpolate complex, non-linear datasets, but they often act as "black boxes," making their predictions challenging to interpret. In order to gain insight into the relationship between glass composition and VHT response, we employed the SHapley Additive exPlanations (SHAP) analysis. This method is based on the coalitional game theory, originally proposed by Lloyd Shapley[54], and it is intended to address a classic math problem—how to attribute an outcome to the presence of multiple contributing factors. Under the ideal conditions assumed by Shapley's theory, the "pure" contribution of each factor (namely, marginal contribution, which is presented by the Shapley value in SHAP analysis) can be determined by analyzing a large group of instances that record variations in these factors and the corresponding changes in the outcome. Importantly, this method, although also permutation-based, differs from conventional permutation feature importance analysis, which focuses on the reduction in prediction accuracy, as the Shapley value directly reflects the actual influence on the same scale of the outcome.

More recently, Shapley's theory has been incorporated into the SHAP analysis for interpreting machine learning models. In brief, this analysis estimates the marginal contributions of the model's input features to the output[20,21]. A significant advantage of the SHAP analysis is its ability to reliably approximate the Shapley value without requiring high computational demands for calculating the exact value. This makes it feasible to apply this approach to large datasets and sophisticated machine learning models. Over the past years, SHAP analysis has successfully been utilized to interrogate various machine learning models in several glass-related studies[44,55–57]. The SHAP analysis involves calculating outcome differences by permuting the factor of interest (between its actual value and its mean value) across all instances under different conditions. By averaging these differences over all instances, we can assess the marginal contribution of the factor of interest (i.e., Shapley value). In our analysis, each glass sample in the VHT dataset corresponds to an "instance", the "factors" are the oxide molar fractions of the glass, and the "outcome" corresponds to the logarithm VHT alteration rate predicted by the model. Note that, depending on the actual influence of an oxide, the Shapley value can be either positive or negative, wherein a positive value signifies an increase in the VHT alteration rate, and vice versa.

References

1. Bergmann, L. M. et al. River Protection Project System Plan, Revision 9 - 21183. WM2021: 47 Annual Waste Management Conference, United States (2021).
2. Marcial, J., Riley, B. J., Kruger, A. A., Lonergan, C. E. & Vienna, J. D. Hanford low-activity waste vitrification: a review. *J. Hazard. Mater.* **461**, 132437 (2024).
3. Xu, X. et al. Machine learning enabled models to predict sulfur solubility in nuclear waste glasses. *ACS Appl. Mater. Interfaces* **13**, 53375–53387 (2021).
4. Pierce, E. M. et al. Laboratory testing of bulk vitrified low-activity waste forms to support the 2005 integrated disposal facility performance assessment. https://www.osti.gov/biblio/15020690-laboratory-testing-bulk-vitrified-low-activity-waste-forms-support-integrated-disposal-facility-performance-assessment. https://doi.org/10.2172/15020690 (2005).
5. ASTM C1285-21. *Standard test methods for determining chemical durability of nuclear, hazardous, and mixed waste glasses and multiphase glass ceramics: The Product Consistency Test (PCT)*. (ASTM International, West Conshohocken, PA, 2021).
6. ASTM C1663-18. *Standard test method for measuring waste glass or glass ceramic durability by vapor hydration test*. (ASTM International, West Conshohocken, PA, 2018).
7. DOE. Design, Construction, and Commissioning of the Hanford Tank Waste Treatment and Immobilization Plant, Contract DE-AC27-01RV14136, as amended, U.S. Department of Energy, Office of River Protection, Richland, WA (2000).
8. Jiřička, A., Vienna, J. D., Hrma, P. & Strachan, D. M. The effect of experimental conditions and evaluation techniques on the alteration of low activity glasses by vapor hydration. *J. Non-Cryst. Solids* **292**, 25–43 (2001).
9. Reiser, J. T. et al. *Product Consistency Test and Vapor Hydration Test Comparisons of a Radioactive Hanford Waste Glass with its Non-Radioactive Simulant Glass*, PNNL-34123, Rev. 1. https://www.osti.gov/biblio/1998870. https://doi.org/10.2172/1998870 (2023).
10. Kruger, A. A., Kim, D.-S. & Vienna, J. D. *Preliminary ILAW formulation algorithm description, 24590 LAW RPT-RT-04-0003, Rev. 1*. https://www.osti.gov/biblio/1110191-preliminary-ilaw-formulation-algorithm-description-law-rpt-rt-rev. https://doi.org/10.2172/1110191 (2013).
11. Vienna, J. D. Compositional models of glass/melt properties and their use for glass formulation. *Procedia Mater. Sci.* **7**, 148–155 (2014).
12. Vienna, J. D. et al. *Hanford immobilized LAW product acceptance testing: tanks focus area results*. https://www.osti.gov/biblio/15001162. https://doi.org/10.2172/15001162 (2001).
13. Piepel, G. F. et al. *ILAW PCT, VHT, Viscosity, and Electrical Conductivity Model Development: VSL-07R1230-1*, ORP-56502. https://www.osti.gov/biblio/1110826. https://doi.org/10.2172/1110826 (2013).
14. Vienna, J. D. et al. *2016 Update of Hanford Glass Property Models and Constraints for Use in Estimating the Glass Mass to Be Produced at Hanford by Implementing Current Enhanced Glass Formulation Efforts*. https://www.osti.gov/biblio/1772236-update-hanford-glass-property-models-constraints-use-estimating-glass-mass-

produced-hanford-implementing-current-enhanced-glass-formulation-efforts. https://doi.org/10.2172/1772236 (2016).

15. Vienna, J. D., Kim, D.-S., Skorski, D. C. & Matyas, J. *Glass Property Models and Constraints for Estimating the Glass to Be Produced at Hanford by Implementing Current Advanced Glass Formulation Efforts*. https://www.osti.gov/biblio/1170502-glass-property-models-constraints-estimating-glass-produced-hanford-implementing-current-advanced-glass-formulation-efforts. https://doi.org/10.2172/1170502 (2013).

16. Vienna, J. D. et al. *Glass Property-Composition Models for Support of Hanford WTP LAW Facility Operation*, PNNL-30932 Rev. 2. https://www.osti.gov/biblio/1862823. https://doi.org/10.2172/1862823 (2022).

17. Song, Y., Wang, Y., Wang, K., Sant, G. & Bauchy, M. Decoding the genome of cement by Gaussian Process Regression. In *Proc. of the Conference on Neural Information Processing Systems.* Workshop on Machine Learning for Engineering Modeling, Simulation and Design (2020).

18. Piepel, G. F., Cooley, S. K., Vienna, J. D. & Crum, J. V. *Experimental Design for Hanford Low-Activity Waste Glasses with High Waste Loading*. https://www.osti.gov/biblio/1212244. https://doi.org/10.2172/1212244 (2015).

19. Bishnoi, S. et al. Predicting Young's modulus of oxide glasses with sparse datasets using machine learning. *J. Non-Cryst. Solids* **524**, 119643 (2019).

20. Shapley, L. S. & Roth, A. E. *The Shapley Value: Essays in Honor of Lloyd S. Shapley* (Cambridge University Press, 1988).

21. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems 30* (eds. Guyon, I. et al.) 4765–4774 (Curran Associates, Inc., 2017).

22. Ouyang, B. et al. Using machine learning to predict concrete's strength: learning from small datasets. *Eng. Res. Express* **3**, 015022 (2021).

23. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).

24. Kumar, N. et al. Machine learning constrained with dimensional analysis and scaling laws: simple, transferable, and interpretable models of materials from small datasets. *Chem. Mater.* **31**, 314–321 (2019).

25. Zhang, Y. & Ling, C. A strategy to apply machine learning to small datasets in materials science. *npj Comput. Mater.* **4**, 1–8 (2018).

26. Mauro, J. C. Topological constraint theory of glass. *Am. Ceram. Soc. Bull.* **90**, 7 (2011).

27. Bauchy, M. Deciphering the atomic genome of glasses by topological constraint theory and molecular dynamics: a review. *Comput. Mater. Sci.* **159**, 95–102 (2019).

28. Phillips, J. C. Topology of covalent non-crystalline solids I: short-range order in chalcogenide alloys. *J. Non-Cryst Solids* **34**, 153–181 (1979).

29. Lu, Z. P. & Liu, C. T. Glass formation criterion for various glass-forming systems. *Phys. Rev. Lett.* **91**, 115505 (2003).

30. Sun, K.-H. Fundamental condition of glass formation*. *J. Am. Ceram. Soc.* **30**, 277–281 (1947).

31. Varshneya, A. K. *Fundamentals of Inorganic Glasses* (Elsevier, 2013).

32. Vienna, J. D., Neeway, J. J., Ryan, J. V. & Kerisit, S. N. Impacts of glass composition, pH, and temperature on glass forward dissolution rate. *npj Mater. Degrad.* **2**, 1–12 (2018).

33. Abraitis, P. K., McGrail, B. P. & Trivedi, D. P. The effects of silicic acid, aluminate ion activity and hydrosilicate gel development on the dissolution rate of a simulated British magnox waste glass. *Sci. Basis Nuclear Waste Manage. XXII* **556**, 401–408 (1999).

34. Frankel, G. et al. Recent advances in corrosion science applicable to disposal of high-level nuclear waste. *Chem. Rev.* **121**, 12327–12383 (2021).

35. Zhen-Wu, B. Y. et al. Predicting zeolites' stability during the corrosion of nuclear waste immobilization glasses: comparison with glass corrosion experiments. *J. Nucl. Mater.* **547**, 152813 (2021).

36. Russell, R. L. et al. *Enhanced Hanford Low-Activity Waste Glass Property Data Development: Phase 2*. https://www.osti.gov/biblio/1813429-enhanced-hanford-low-activity-waste-glass-property-data-development-phase. https://doi.org/10.2172/1813429 (2021).

37. Rickman, J., Lookman, T. & Kalinin, S. Materials informatics: from the atomic-level to the continuum. *Acta Mater.* **168**, 473–510 (2019).

38. Piepel, G. F., Szychowski, J. M. & Loeppky, J. L. Augmenting Scheffé linear mixture models with squared and/or crossproduct terms. *J. Qual. Technol.* **34**, 297–314 (2002).

39. Seabold, S. & Perktold, J. Econometric and Statistical Modeling with Python. https://doi.org/10.25080/MAJORA-92BF1922-011 (2010).

40. Heaton, J. *Introduction to Neural Networks with Java*. (Heaton Research, Inc., 2008).

41. Abiodun, O. I. et al. State-of-the-art in artificial neural network applications: a survey. *Heliyon* **4**, e00938 (2018).

42. Cassar, D. R. GlassNet: a multitask deep neural network for predicting many glass properties. *Ceram. Int.* **49**, 36013–36024 (2023).

43. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems* vol. 32 (Curran Associates, Inc., 2019).

44. Song, Y. et al. Machine learning enables rapid screening of reactive fly ashes based on their network topology. *ACS Sustain. Chem. Eng.* **9**, 2639–2650 (2021).

45. Ouyang, B., Song, Y., Li, Y., Sant, G. & Bauchy, M. EBOD: an ensemble-based outlier detection algorithm for noisy datasets. *Knowl.-Based Syst.* **231**, 107400 (2021).

46. Müller, A. C. & Guido, S. *Introduction to Machine Learning with Python: A Guide for Data Scientists* (O'Reilly Media, Inc., 2016).

47. Carroll, R. J., Gutierrez, R. G., Wang, C. Y. & Wang, S. Local linear regression for generalized linear models with missing data. *Ann. Stat.* **26**, 1028–1050 (1998).

48. Schulz, E., Speekenbrink, M. & Krause, A. A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions. *J. Math. Psychol.* **85**, 1–16 (2018).

49. MacKay, D. J. C., Kay, D. J. C. M. & MacKay vid J. C. *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, 2003).

50. Gardner, J., Pleiss, G., Weinberger, K. Q., Bindel, D. & Wilson, A. G. GPyTorch: blackbox matrix–matrix gaussian process inference with GPU acceleration. *Advances in Neural Information Processing Systems.* Vol. 31 (Curran Associates, Inc., 2018).

51. Duvenaud, D. Automatic Model Construction with Gaussian Processes (University of Cambridge, 2014) https://doi.org/10.17863/CAM.14087.

52. Jablonka, K. M., Ongari, D., Moosavi, S. M. & Smit, B. Big-data science in porous materials: materials genomics and machine learning. *Chem. Rev.* **120**, 8066–8129 (2020).

53. Christensen, R. *Analysis of Variance, Design, and Regression: Applied Statistical Methods* (CRC Press, 1996).

54. Shapley, L. S. Quota solutions of N-person games. *Contrib. Theory Games* **2**, 343–359 (1953). 1952-04-07.

55. Ravinder et al. Artificial intelligence and machine learning in glass science and technology: 21 challenges for the 21st century. *Int. J. Appl. Glass Sci.* **12**, 277–292 (2021).

56. Cassar, D. R. et al. Predicting and interpreting oxide glass properties by machine learning using large datasets. *Ceram. Int.* **47**, 23958–23972 (2021).

57. Zaki, M., Jayadeva & Krishnan, N. M. A. Extracting processing and testing parameters from materials science literature for improved property prediction of glasses. *Chem. Eng. Process. Process Intensif.* **180**, 108607 (2022).

58. Feng, X., Hrma, P. R. & Westsik, J. *Glass Optimization for Vitrification of Hanford Site Low-Level Tank Waste*. https://www.osti.gov/biblio/219300. https://doi.org/10.2172/219300 (1996).
59. Vienna, J. D. et al. *Hanford Immobilized LAW Product Acceptance: Initial Tanks Focus Area Testing Data Package*. https://www.osti.gov/biblio/965231. https://doi.org/10.2172/965231 (2001).

Author contributions
Y.S., X.L., K.W., and J.D.V. were responsible for data processing and modeling analysis. Y.S., X.L., and J.D.V. drafted the manuscript. J.D.V., J.V.R., M.S.S., and M.B. reviewed the manuscript. J.D.V. and M.B. were responsible for the conception.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41529-024-00458-6.

Correspondence and requests for materials should be addressed to John D. Vienna or Mathieu Bauchy.

Reprints and permissions information is available at http://www.nature.com/reprints

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