Enabling extrapolation of Young's modulus of $CaO-Al_2O_3-SiO_2$ ternary glasses by topology-informed machine learning

Yang, Kai; Song, Yu; Li, Yuhai; Smedskjaer, Morten M.; Bauchy, Mathieu; Rosner, Fabian

Published in:
Journal of Non-Crystalline Solids

DOI (link to publication from Publisher):
10.1016/j.jnoncrysol.2025.123610

Publication date:
2025

Document Version
Publisher's PDF, also known as Version of record

Journal of Non-Crystalline Solids 666 (2025) 123610

Contents lists available at ScienceDirect

Journal of Non-Crystalline Solids

journal homepage: www.elsevier.com/locate/jnoncrysol

Enabling extrapolation of Young's modulus of CaO-Al_2O_3-SiO_2 ternary glasses by topology-informed machine learning

Kai Yang [a,b], Yu Song [a], Yuhai Li [a], Morten M. Smedskjaer [c], Mathieu Bauchy [a], Fabian Rosner [a,b,d,*]

[a] Department of Civil and Environmental Engineering, University of California, Los Angeles, CA 90095, USA
[b] Renewable Energy and Chemical Technologies (REACT) Lab, Department of Civil and Environmental Engineering, University of California, Los Angeles, CA 90095, USA
[c] Department of Chemistry and Bioscience, Aalborg University, 9220 Aalborg, Denmark
[d] Institute for Carbon Management (ICM), University of California, Los Angeles, CA 90095, USA

ARTICLE INFO

Keywords:
Glass
Mechanical properties
Topological constraint theory
Machine learning
Molecular dynamics

ABSTRACT

The application of machine learning (ML) in material discovery, particularly in the design of novel materials like glasses, has shown considerable promise. However, the efficacy of data-driven ML approaches is often hindered by the limited volume and representativeness of material datasets. While these approaches demonstrate notable success in interpolating data, they tend to perform inadequately in extrapolation tasks, which are crucial in the context of material discovery. In this study, we address this challenge by incorporating topological knowledge, derived from the atomic structures of glasses, to inform ML models with physics-based insights. To showcase this approach, we focus on predicting Young's modulus of CaO-Al_2O_3-SiO_2 glasses. By leveraging the topological information, i.e., the fractions of bond-stretching and bond-bending constraints, we transform a non-linear composition-property mapping to a higher-linearity topology-property mapping to improve the extrapolation abilities of ML models. Our results demonstrate that the topology-informed ML approach maintains comparable prediction accuracy within the training domain while significantly improving performance in extrapolating the Young's modulus of glasses beyond the training domain. Therefore, our topology-informed approach can offer a more efficient and expedited pathway towards the discovery of new glass materials in unexplored domains.

1. Introduction

Over the past decade, material scientists have extensively employed machine learning (ML) techniques to advance their knowledge across various domains including battery materials, alloys, and ceramics [1–3]. In particular, the field of glass science has benefited greatly from the advances in ML, including the development of cover glasses for electronic devices, optical fibers, bioactive glasses, and glasses for nuclear waste immobilization [4–8]. Among all the promising applications of ML in glass science, predictive models that map glass composition to material properties have drawn great attention [9–13]. Such composition-property mappings are of significance in the discovery of novel glass compositions with desired properties, for a range of applications [14,15].

Although many published predictive models for oxide glasses have demonstrated good results in terms of accuracy metrics, such as coefficient of determination (R^2) or root mean square error (RMSE), many of these models are limited to data interpolation, meaning that the model itself is built to predict what it has already known [16]. Traditional machine learning workflows involve randomly selecting a test set from all available data and keeping it unknown to the model during training. However, this selected test set is part of the training domain, which hinders the evaluation of the model's extrapolation ability. This is problematic since the discovery of new high-performance glasses requires researchers to extrapolate from the current knowledge base of glasses and their properties —in other words; the test-set domain. Such an extrapolation task is intrinsically challenging for traditional ML approaches, which are solely conditioned by the training data.

Over the years, several approaches have been developed to enhance the extrapolation abilities of ML models for materials discovery with the goal to provide better training procedures. For example, Meredig et al. proposed a novel standardized technique, known as leave-one-cluster-out (LOCO) cross-validation (CV), as a replacement for the conventional k-fold CV for assessing model performance [16]. The traditional

* Corresponding author at: 5731-H Boelter Hall, Los Angeles, CA 90095-1593, USA.
E-mail address: fabianrosner@g.ucla.edu (F. Rosner).

https://doi.org/10.1016/j.jnoncrysol.2025.123610
Received 28 February 2025; Received in revised form 2 May 2025; Accepted 6 May 2025
Available online 2 June 2025

k-fold CV method randomly shuffles and splits data into a training-set and a test-set, which makes it inadequate for measuring the extrapolation ability of ML algorithms. By contrast, the LOCO—CV approach uses an unsupervised clustering algorithm to separate the dataset based on the proximity of the data points to one another within the material domain. The test-set, which is the cluster left out, can then be used to evaluate the extrapolation performance of the ML algorithms.

Another approach is based on feature engineering, which involves training ML models on gross-level property-based descriptors [3]. Ghiringhelli et al. showed that mathematically computed and selected descriptors can play a critical role in extrapolation and material discovery [17,18]. This method first identifies the potentially relevant primary features (e.g., atom band gap or atomic density), and builds descriptors with non-linear prototype functions, such as 1/x or ln(x). Next, thousands or millions of new compound descriptors can be created by forming algebraic combinations (e.g., ln(x)/x or x^2e^{-x}). The least absolute shrinkage and selection operator (LASSO) is then applied to select the most statistically correlated features for the standard ML workflow. The LASSO regularization technique effectively reduces model complexity by ignoring unimportant features through L1 regularization [19], thereby mitigating the risk of overfitting in high-dimensional feature spaces. The gross-level feature engineering approach achieves its objective through two key factors: (1) it converts the extrapolation challenge into an interpolation problem using new feature inputs, and (2) it selects the most important features from tens of thousands of new features. Their work reveals that a key to improving ML model extrapolation ability lies in developing relevant descriptors that shows high degree of linearity with target properties without losing information.

Although both approaches offer improvements in extrapolation abilities, they are subject to limitations. LOCO—CV necessitates a large dataset that can be partitioned into distinct clusters, which poses a challenge as available materials datasets are often imbalanced and insufficient in size, hindering the formation of well-clustered data for model training and validation. On the other hand, the performance of feature engineering heavily relies on the selection of primary features and mathematical functions. Moreover, as the computed features are not specifically targeted or designed with the underlying physics in mind, but rather generated in a non-discriminative manner, a large number of features are typically required to train a machine learning model to achieve satisfactory performance. This expansion of dimensionality can lead to the notorious 'curse of dimensionality' [20], increasing the risk of overfitting, where the model learns the noise pattern present in the new features rather than capturing the underlying physical phenomena.

Among the recent knowledge advances in glass science, topological constraint theory (TCT) offers a direct and well-studied route that can provide high-linearity relationships to bridge glass compositions and their material properties [21–25]. TCT simplifies complex disordered atomic networks into simpler trusses and nodes, the rigidity of which are determined by the number of constraints. There are two types of topological constraints, namely, (1) the radial 2-body bond stretching (BS) constraints that maintain the averaged bond length and (2) the angular 3-body bond-bending (BB) constraints that fix the averaged interatomic angles. In particular, TCT excels at capturing the key connectivity from the glass atomic network, while filtering out less relevant structural details that, for most properties, only have second-order impacts on the macroscopic properties [26,27]. Han et al. proposed topology-informed machine learning approach to predict silicate glass dissolution kinetics, tackling shortcomings of conventional ML techniques [28]. By integrating structural information about the glass network topology, their models achieved greater accuracy, reduced complexity, and improved extrapolation to unfamiliar glass compositions. While this method advances beyond traditional approaches, it still relies on interpolation within a transformed input space, where compositional data is converted to topological descriptors. This transformation process bears similarities to using gross-level property-based parameters, representing an intermediate step between purely compositional models and fully predictive physical frameworks. Other works have also shown that TCT enables the development of a variety of analytical models to predict glass properties [24–30], as well as properties of other disordered materials, such as cementitious materials and fly ashes [31,32].

In this study, we propose a topology-informed ML approach that focuses on extrapolating the Young's modulus of calcium aluminosilicate (CAS) glasses using topological features. We compute the topological features using an analytical model derived from our previous work [33] and introduce them as new inputs to train ML algorithms, namely, polynomial regression (PR), random forest (RF), and multi-layer perceptron (MLP). Our results demonstrate that the proposed topology-informed MLP model yields accurate predictions for both interpolation and extrapolation tasks, outperforming all composition-informed models. More importantly, we show that our ML models extrapolate successfully with computed topological input space and yield good predictions on extrapolated compositional inputs. We discuss that developing highly linearized physics-based features that correlate with the target properties can serve as an alternative and explicit solution to enable extrapolation capabilities in ML applications. This work highlights the importance of integrating physics-based insights into the ML framework to address the challenges associated with extrapolation in materials discovery.

2. Methods

2.1. Data collection

All Young's modulus data were computed from molecular dynamics (MD) simulations in our previous work [9]. Note that, each reported Young's modulus value represents an average calculation from six independent simulations for each composition. Further details regarding the simulation methodology can be found in the reference [9] and the Supplementary Material S1.

2.2. Analytical model for topology constraints enumeration

To compute the topological features, a fully analytical model from our previous work was adopted to enumerate the number of constraints of CAS glasses based on their compositions [33]. Given the concentration of each oxide constituent in a CAS glass, this analytical model estimates the numbers of featured atom species, including Ca atoms, Si atoms, 4-fold or 5-fold Al atoms, and free oxygen (FO), non-bridging oxygen (NBO), bridging oxygen (BO) and tri-cluster oxygen (TO)

Table 1
Summary of the bond-stretching (BS) and bond-bending (BB) constraints contributed by each atom species in calcium aluminosilicate glasses. For Al and O atoms, the constraints are differentiated by their coordination numbers. For Ca atoms, the constraints depend on the type of O atoms to which they are connected.

	Glassy state	
Species	BS	BB
Si atoms	4	5
Al atoms		
4-fold Al	4	5
5-fold Al	5	0
Ca–O bonds		
CaFO	1	/
Ca–NBO	1	/
Ca–BO	0	/
Ca–TO	0	/
O atoms		
FO	/	0
NBO	/	0
BO	/	1
TO	/	3

atoms. Readers are referred to the previous study for details of the enumeration of those atoms [33]. Table 1 shows the numbers of bond-stretching (BS) and bond-bending (BB) constraints associated with each of the atom species. Note that for Ca atoms, we counted the number of constraints based on the chemical bond formed by Ca and different types of oxygen atoms. The total number of BS and BB constraints was the summation of constraints contributed by each atom species. After the enumeration of BS and BB constraints in the system, we used a volumetric model to compute the volume of glass samples (see Supplementary Material S2). Then, the two topological inputs n_C / V and BS/n_C were computed for all CAS glasses and used to train ML models.

2.3. Machine learning models and workflow

Machine learning algorithms, such as polynomial regression (PR), random forest (RF), or multi-layer perceptron (MLP), have been successfully applied in predicting different properties of oxide glasses [9–35]. In particular, MLP is a feedforward neural network that consists of input layer, output layer and hidden layers, with each layer fully connected to the next [36]. Each layer in MLP has some nodes and each node is a simple mathematical function that takes the weighted sum of its inputs, adds a bias term, and applies an activation function to produce an output. The weights and biases of the nodes are the trainable parameters of the model, which are adjusted during the training process to minimize the error between the predicted output and the actual output.

Here, all algorithms are imported from the Scikit learn package [37] and follow the standard ML workflow, including data normalization, data splitting, feature engineering, and hyperparameter tuning. We firstly applied min-max normalization to inputs, and then split dataset based on interpolation or extrapolation purposes. To set a fair competition, all interpolation models randomly choose 50 % data points to train and test on the remaining. For extrapolation models, the splitting conditions are listed in the Table 4, with a train-test split ratio close to 50–50. We used 5-fold cross validation to tune the hyperparameters (i.e., the degree of polynomial, the number of trees and the number of neurons) for all models to avoid underfitting or overfitting. The details of selecting hyperparameters for ML models can be found in the Supplementary Material S5 and S7. The composition-informed models used glass compositions as inputs, while the topology-informed models used computed topological features as inputs. Finally, we computed the root mean square error (RMSE) to quantify the performance of each model on training and test set.

2.4. Model interpretation

SHAP (SHapley Additive exPlanations) is a technique for interpreting the output of machine learning models [38], based on the concept of Shapley values from cooperative game theory [39]. We employ the SHAP Python module to interpret predictions generated by the trained models. In SHAP, the contribution of each feature on the model output is allocated based on their marginal contribution, and a SHAP summary plot can be visualized to display the magnitude and direction of each feature's impact on the model's output.

3. Results

3.1. Comparison between composition-informed and topology-informed inputs

We first compare the training processes of the composition-informed approach and the topology-informed approach used in our work. As illustrated in Fig. 1a, the conventional ML approach typically aims to capture the complex, non-linear relationships between compositional inputs and the target properties [9–15]. Conversely, our topology-informed ML approach establishes a composition-structure-property mapping, whereby the compositional inputs are first transformed into topological features by leveraging the underlying physics. Subsequently, we train an ML model to capture the mapping between the topological inputs and the target properties. It is important to note that we do not merely append the new features as supplementary inputs for ML models; instead, we completely replace the compositional inputs with the topological features. Unlike the gross-level feature engineering approach [17], our methodology thus circumvents the expansion of feature dimensionality, as the computed topological features inherently encapsulate glass compositional information (see Methods).

We use simulated data points as benchmark in this work to evaluate

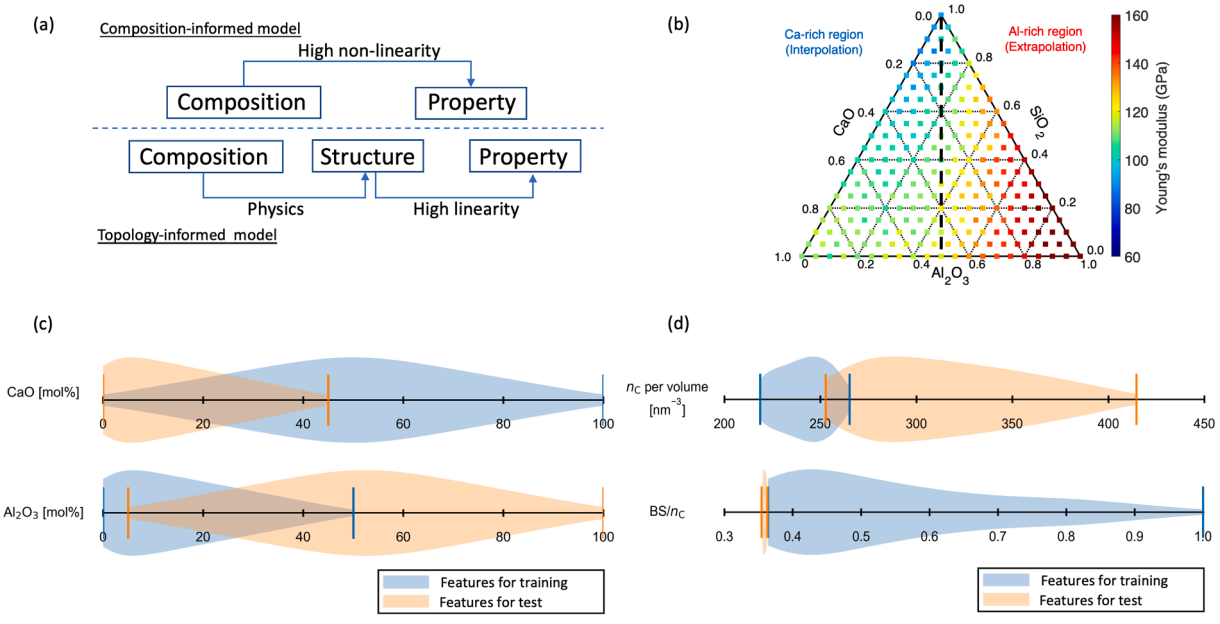

Fig. 1. Model input preparation and comparison. (a) Comparison between composition-informed and topology-informed ML training approaches. (b) Simulated Young's modulus for calcium aluminosilicate (CAS) glasses over the entire domain. (c) Distribution of inputs for the training and test set of composition-informed extrapolation model. (d) Distribution of inputs for the training and test set of topology-informed extrapolation model.

the ability of interpolation and extrapolation using composition-informed and topology-informed ML approaches. In detail, we simulate 231 different glass compositions using MD simulations and compute their Young's modulus. These glass compositions are homogeneously distributed throughout the entire CAS domain, with a 5 mol% increment in each CaO, SiO_2, Al_2O_3 constituent. For each composition, we conduct six independent simulations, and an average Young's modulus, E, was computed thereafter [9]. Fig 1b presents the variation of the averaged Young's modulus across the investigated CAS glasses over the entire ternary space, wherein each composition is color-coded based on the magnitude of Young's modulus. Two evident trends are noted from this ternary plot: (1) along the vertical direction, E gradually increases as the concentration of [SiO_2] becomes lower, and (2) along the horizontal direction, E broadly increases with the rise of the concentration of [Al_2O_3], which echo with results from experiments [40].

To set up an extrapolation task, we divide the CAS dataset into two distinct regions: (1) the Ca-rich region and (2) the Al-rich region, corresponding to the left and right halves of the CAS ternary plot, respectively, as shown in Fig. 1b Specifically, we limit the training data to the data points residing within the Ca-rich region, and the data points located in the Al-rich region are reserved as a test set to evaluate the performance of each model's extrapolation ability. This splitting method gives ML algorithms a more challenging task to extrapolate since the algorithms need to capture different increasing slopes of Young's modulus in the two regions [25], as it will be highlighted below. As Al atoms replace Ca atoms, the Young's modulus increases, with a notable change in growth rate at the so-called charge-compensation line ([Al_2O_3] = [CaO]). This transition is attributed to structural changes in the glass network: in Ca-rich regions, Al atoms form tetrahedral structures compensated by Ca atoms, while in Al-rich regions, they increase polymerization by forming over-coordinated 5-fold and 6-fold Al units due to the deficit of Ca cations, resulting in a more stress-rigid network [41–44]. More details can be found in Ref. [33]. This structural transition, manifesting as different Young's modulus development rates in the two regions, complicates extrapolation for composition-informed models due to the absence of explicit structural information. Furthermore, we also use other splitting methods as listed in Table 4.

We emphasize that there are only 2 inputs used for both the composition-informed and topology-informed approaches. Fig 1c shows that CaO and Al_2O_3 in mol% are the inputs used for the composition-informed ML models, as the molar fraction of SiO_2 is just a function of [CaO] and [Al_2O_3] (i.e., [SiO_2] = 100 % – [CaO] – [Al_2O_3]). For the topology-informed approach, we compute two topological features: (1) the volumetric density of total constraints (n_C/V), and (2) the fraction of bond-stretching constraints in total constraints enumeration (BS/n_C) as inputs. The rationale behind the selection of these features and the methodology for their computation will be discussed in the subsequent sections.

The violin plots in Fig. 1c and d illustrate the distribution of input features for both the training and test sets, comparing composition-informed and topology-informed extrapolation models, respectively. The shape of the violin plot shows the distribution of the data, wherein a broader violin represents a higher data density at the value range and a narrower area represents lower density. It is important to highlight here that our approach differs from the traditional feature engineering method, which uses numerical descriptors to transform an exploration problem into an interpolation problem. In contrast, our topological features maintain a clear separation between the interpolation and extrapolation regions. Notably, our topological inputs exhibit an even more isolated state when comparing with the compositional inputs. That is, our topology-informed training approach still encounters an extrapolation challenge on the topological features when predicting the Young's modulus in the unknown Al-rich region. Additionally, we present the color-coded ternary plots of compositional inputs (CaO and Al_2O_3) and topological inputs (n_C/V and BS/n_C) in Fig. S1. The topological input plots reveal distinct color patterns between Ca-rich and Al-rich regions, highlighting that our model still relies on extrapolation of topological inputs across these compositional domains.

3.2. Composition-informed machine learning approach

3.2.1. Interpolation performance

We first adopt three different ML algorithms used in our previous work [9], namely, polynomial regression (PR), random forest (RF), and multilayer perceptron (MLP), to interpolate the Young's modulus of CAS glasses across the entire compositional space. Note that only the compositions of CaO and Al_2O_3 are used as input features as discussed in the previous section. We randomly split all data points into a training set and a test set with a 50–50 split ratio for the conventional approach. Although this is not a commonly used split ratio, we use the same ratio for extrapolation models in the next section to have a fair competition between models' interpolation and extrapolation ability. All models are trained with the standard ML process, including normalization and 5-fold cross validation for hyperparameter tuning (e.g., degree of polynomial, number of trees, number of neurons etc.).

Table 2 includes the interpolation performance with coefficient of determination (R^2) and root mean square error (RMSE) of training and test set for all algorithms used herein. We observe that all algorithms can well capture the composition-property mapping for Young's modulus of CAS glasses for an interpolation case, despite using only half of the dataset for training. In more details, we present the interpolation results from composition-informed MLP model in Fig. 2, while the performance of PR and RF models can be found in Supplementary Material S4. As shown in Fig. 2a, the MLP model offers good predictions when interpolating Young's modulus of CAS glasses, achieving a 3.43 GPa RMSE on the test set. Fig 2b presents the predicted Young's modulus of CAS glass series at [SiO_2] = 30 mol%, showing that the composition-informed interpolation model gives accurate predictions and well captures the increasing trend of E when the molar difference of [Al_2O_3] – [CaO] increases, particularly the change of slopes for glasses in Ca-rich and Al-rich regions. Moreover, the confidence interval is shown by calculating the standard deviation of predictions from six independently trained models, revealing low uncertainty and good stability of model performance. The trained MLP model in Fig. 2c captures the non-linear mapping between glass compositions and their Young's modulus, providing accurate and continuous predictions of Young's modulus covering the entire CAS domain.

3.2.2. Extrapolation performance

Next, we investigate the performance of extrapolation for the composition-informed ML models using the same ML algorithms. We use the Ca-rich region to train the model and extrapolate on the Al-rich region as discussed in the previous section and keep the train-test split ratio close to the previous section. All models are trained under the same standard processes, including normalization, cross-validation and hyperparameter tune. Details of hyperparameter tuning can be found in the Supplementary Material S5 for all composition-informed models. Table 2 summarizes the R^2 and RMSE of training and test set for extrapolation models with composition inputs. Comparing to model performance for interpolation, we note that although extrapolation models yield good prediction for the training set, all of them fail to extrapolate, giving negative R^2 and extreme high RMSE when extrapolating Young's modulus in the Al-rich region.

Fig 3 presents detailed results predicted by the three composition-informed extrapolation models. Although all models successfully interpolate Young's modulus in the known Ca-rich region (blue points and left-half of ternary plots in Fig. 3), they struggle to accurately extrapolate Young's modulus in the unknown Al-rich region (red points and right-half of ternary plots in Figs. 3). The polynomial regression model tends to overfit the Young's modulus during training, giving extremely large and unrealistic values due to the high degree of polynomial from the validation process. The random forest model exhibits no ability to

Table 2
Comparison between the interpolation and extrapolation ability of machine learning models with composition inputs for the different algorithms used herein, namely, polynomial regression (PR), random forest (RF), and multilayer perceptron (MLP). The coefficient of determination (R^2) and root mean square error (RMSE) are used to describe the accuracy of predictions for training and test set.

Composition inputs	Interpolation				Extrapolation			
	R^2_train	R^2_test	RMSE_train	RMSE_test	R^2_train	R^2_test	RMSE_train	RMSE_test
PR	0.981	0.971	2.388	2.754	0.907	−2716	2.083	795.0
RF	0.993	0.966	1.385	2.987	0.976	−1.910	1.205	26.017
MLP	0.968	0.961	3.253	3.431	0.813	−1.114	2.953	18.174

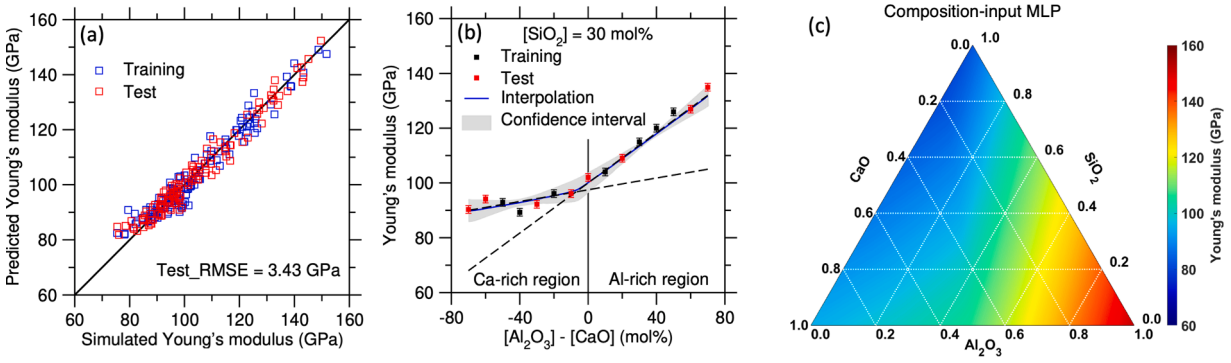

Fig. 2. Performance of the conventional composition-informed model to interpolate Young's modulus of CAS glasses. (a) Comparison between predicted and simulated Young's modulus of CAS glasses using conventional approach. (b) Comparison between simulated (black and red dots) and predicted (blue line) Young's modulus for $[SiO_2] = 30$ mol% glass series using conventional ML model. Solid line represents the charge compensation line ($[Al_2O_3] = [CaO]$) to highlight the Ca-rich and Al-rich regions. Dashed lines are guides for the eye to demonstrate the different slopes in the two regions. (c) Prediction of Young's modulus of CAS glasses from composition-informed MLP model over the entire CAS composition domain.

extrapolate in Al-rich region, producing stepwise and discontinuous predictions, even in the known Ca-rich region (Fig. 3e).

The hyperparameter selection for ML models is conducted under the assumption that the extrapolation (test) set remains unknown, with no prior knowledge of input-output correlations. We evaluate each model's performance on training and validation set (5-fold cross validation) to choose the optimal hyperparameters. As shown in Fig. S3a, the PR model accuracy metrics (R^2 and RMSE) are evaluated against polynomial degree. An 8th-order polynomial demonstrated best performance on both training and validation sets and is subsequently selected for extrapolation testing. However, our domain knowledge suggested that lower-order polynomials can better capture composition-property relationships and avoid overfitting [9]. Therefore, we develop three additional PR models (from 1st to 3rd degree) for comparison in Fig. S3b, c, and d. As expected, all lower-order models exhibit better extrapolation abilities compared to the 8th-degree model. Notably, the 2nd-degree PR model performed well in extrapolating Young's modulus (Fig. S3c) using composition as inputs, although it still underperforms relative to our topology-informed MLP model (Fig. 4h). It is important to emphasize that our selection of the 2nd-degree PR model stemmed from domain expertise rather than from the standard ML workflow that favors the 8th-degree model based solely on the validation results.

In comparison, the MLP model provides the most reasonable predictions; however, it still generates undesirable extrapolations. As shown in Fig. 3h, while the composition-informed MLP model predicts the increased trend of Young's modulus in the Ca-rich region, it does not capture the change of slope of Young's modulus when glasses enter the Al-rich region. Furthermore, we observe that the confidence interval widens as the extrapolations move farther away from the known region, indicating the models' increasing uncertainty regarding predictions in the unexplored domain. The limitations of composition-input extrapolation models demonstrate that traditional composition-property mapping does not offer robust predictions outside the known region, which strictly limit the application of ML in discovering novel glasses that have not yet been synthesized.

3.3. Topology-informed machine learning approach

We now introduce the topology information of CAS glasses as new input features for ML algorithms by replacing the composition inputs. Fig 4a illustrates the general concepts of how to convert the structural information on the atomic level of glasses into topological features. By simplifying atoms and chemical bonds as joints and linkages, we can classify the atomic structures into two basic topological constraints: (1) the radial 2-body bond-stretching (BS) constraints, and (2) the angular 3-body bond-bending (BB) constraints, which fix the inter-atomic distances and angels between atoms, respectively. Here, we enumerate the number of constraints of CAS glasses using an analytical model from our previous work [33] (see Methods), and compute n_C/V and BS/n_C as two new features for the ML models. The first feature n_C/V describes the volumetric density of total constraints in the CAS glasses, wherein the number of total constraints per atom n_C is the summation of BS and BB per atom, and the volume of glasses V can be computed from another analytical model (see Supplementary Material S2). The second feature BS/n_C presents the fraction of BS constraints in total constraints enumeration. Since BS and BB have different weights of contribution towards glasses' stiffness [25–46], the feature BS/n_C provides information about the contributed portion from BS or BB constraints towards the Young's modulus at a specific CAS glass composition.

3.3.1. Interpolation performance

We use the topological inputs to train the same ML algorithms selected previously and access their interpolation ability. Table 3 summarizes the R^2 and RMSE values for the training and test set of PR, RF and MLP for topology-informed interpolation predictions. We note that all algorithms yield reasonable predictions on training and test set. As all R^2 are larger than 0.94 and all RMSE values are smaller than 4 GPa on the test set, we conclude that introducing topological features to ML algorithms does not impair their ability on interpolation, comparing to the results in Table 1. In fact, all algorithms demonstrate proficiency in capturing composition-property or structure-property mappings when

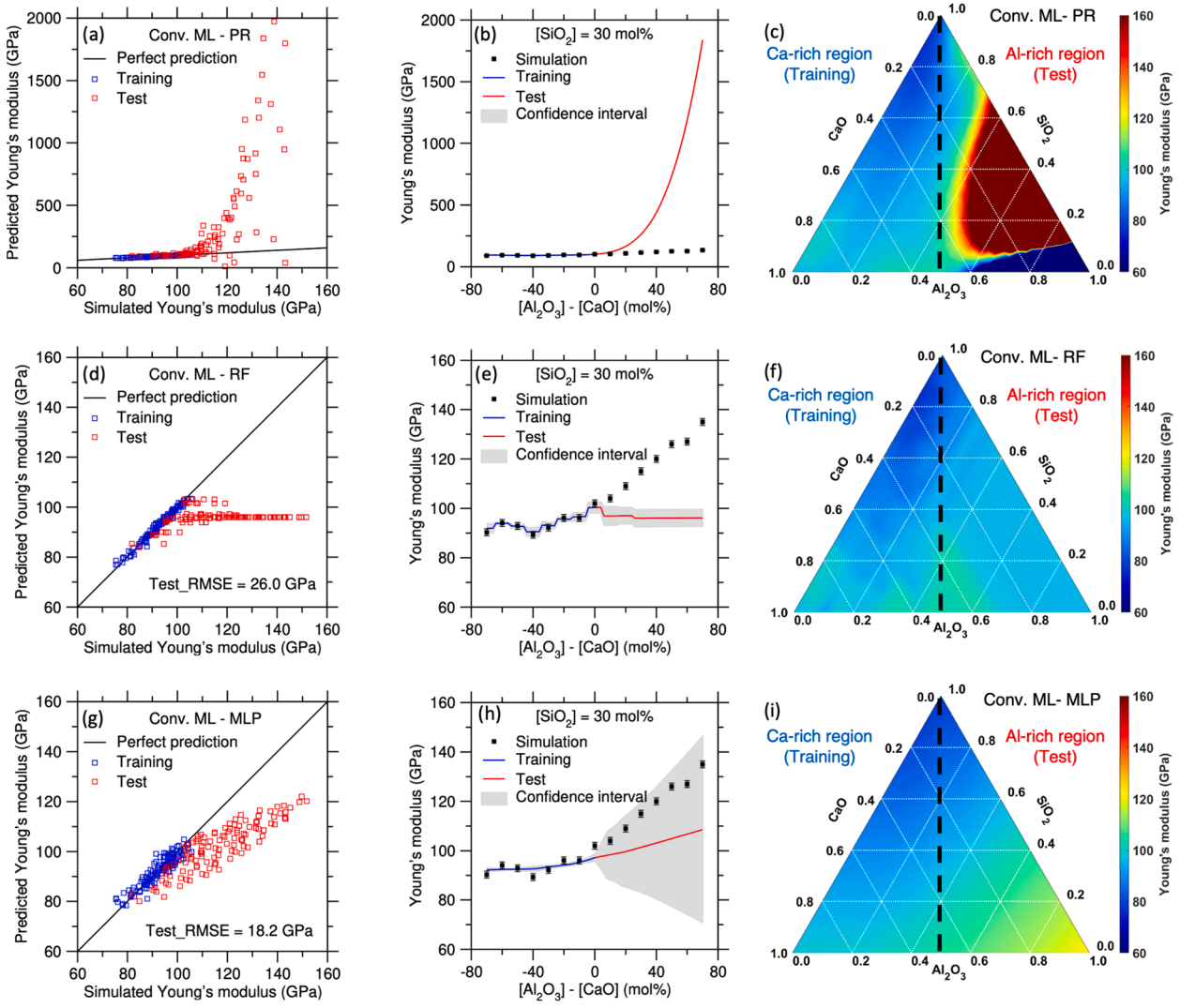

Fig. 3. Extrapolating Young's modulus of CAS glass by conventional machine learning, including polynomial regression (PR), random forest (RF) and multilayer perceptron (MLP) models. (a), (d), (g) Comparison between predicted and simulated Young's modulus of CAS glasses using conventional ML approach for interpolation (training) and extrapolation (test) regions. (b), (e), (h) Comparison of Young's modulus between simulated and predicted results for $[SiO_2] = 30$ mol% glass series in CAS glass family on interpolation and extrapolation regions with conventional ML models. The grey areas are confidence interval. (c), (f), (i) Prediction (interpolating in Ca-rich region and extrapolating in Al-rich region) of Young's modulus of CAS glasses using conventional ML approach over the entire CAS domain. The vertical dash line is the charge-compensate line, where $[Al_2O_3] = [CaO]$, to separate the two regions.

interpolating. Detailed plots that present the predictions of topology-informed interpolation models are shown in Supplementary materials S6.

3.3.2. *Extrapolation performance*

We employ the topological features for PR, RF and MLP algorithms to train in the Ca-rich region and extrapolate in the Al-rich region of CAS glasses, following the same training process. Details of hyperparameter tuning can be found in the Supplementary Material S7 for all topology-informed models. Table 3 collects the performance of topology-informed extrapolation models with their R^2 and RMSE on the training and test set. All plots presenting detailed results for these models are shown in Fig. 4.

Both topology-informed PR and RF models fail to extrapolate the Young's modulus in the Al-rich region. Although they can offer good predictions on the training set, they both give unreasonable extrapolations in terms of R^2 and RMSE on test set as listed in Table 2. The plots shown in Fig. 4 demonstrate that these algorithms are intrinsically unsuitable for extrapolation tasks, wherein the polynomial regression shows a tendency to overfit during the training process (see Fig. 4b, c, d), while the discrete nature of random forest fails to provide reliable extrapolations as it does not have a smooth representation of the input-output relationship (see Fig. 4e, F, g). The topology-informed PR model is optimized at 4th degree, as evidenced by Fig. S5a, which shows this best performance on both training and validation datasets. In addition, we train three additional PR models (from degree 1 to degree 3) for comparison in Fig. S5b, S5c and S5d. Notably, the 1st degree PR model exhibits good extrapolation performance, though it still underperforms than the topology-informed MLP model (Fig. 4h). It should be emphasized that our implementation of the 1st degree PR model is guided by domain knowledge rather than standard machine learning practices, which recommends the 4th-degree PR model based on validation performance metrics.

On the other hand, following the same process, the new topology-informed MLP model yields accurate predictions in both interpolating and extrapolating regions. The RMSE value for the topology-informed extrapolation MLP is 4.65 GPa, demonstrating a substantial improvement compared to the composition-informed extrapolation MLP with an RMSE of 18.2 GPa. Fig 4i provides detailed predictions of the exemplary CAS glasses series at $[SiO_2] = 30$ %. Notably, the topology-informed

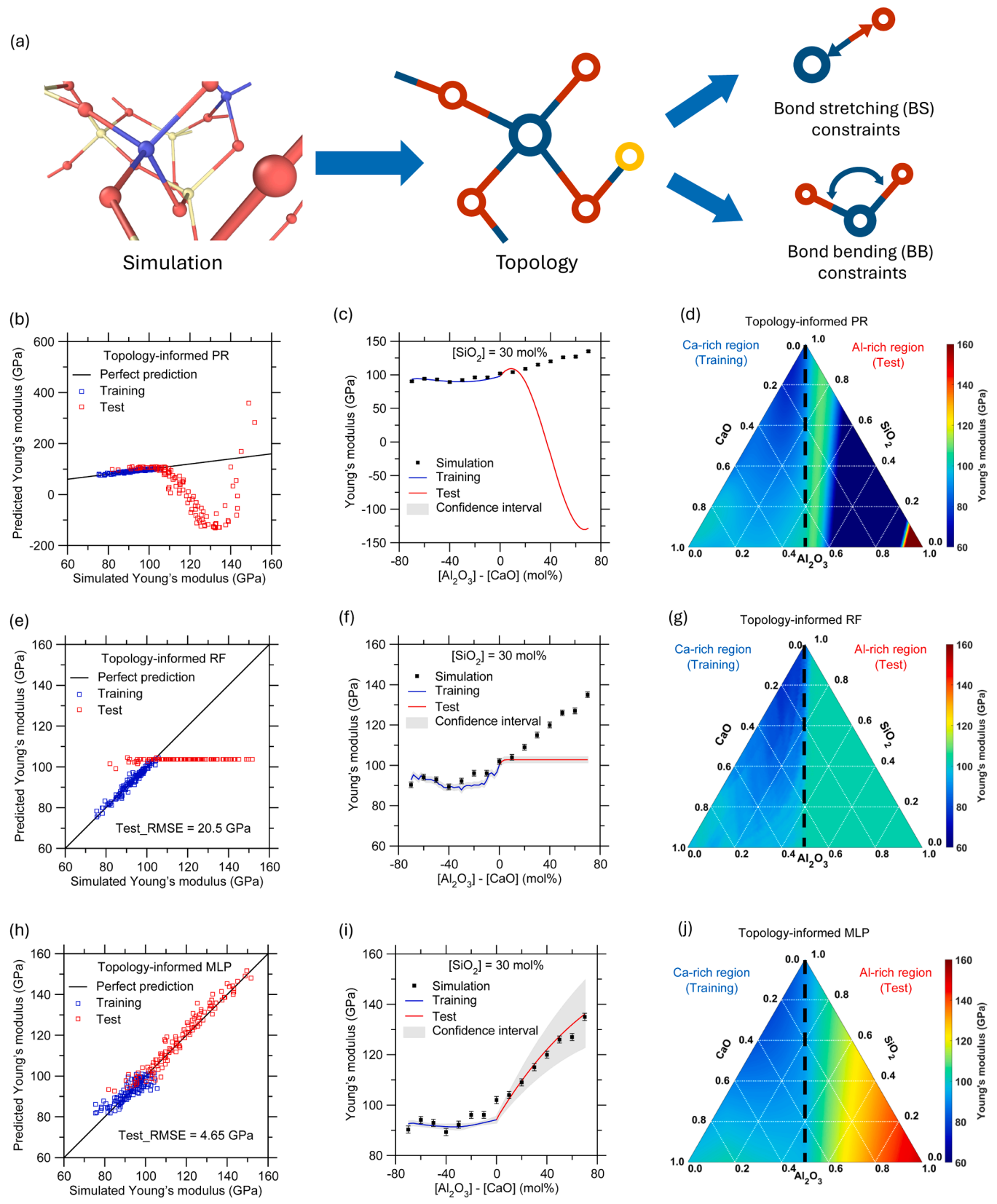

Fig. 4. Topology-informed models extrapolate the Young's modulus of CAS glasses in Al-rich region. (a) Schematic illustration of transforming atomic structures into topological constraints. (b), (e), (h) Comparison between predicted and simulated Young's modulus of CAS glasses using topology-informed approach for interpolation (training) and extrapolation (test) regions. (c), (f), (i) Comparison of Young's modulus between simulated and predicted results for $[SiO_2] = 30$ mol% glass series in CAS glass family on interpolation and extrapolation regions with topology-informed models. The grey areas are confidence interval. (d), (g), (j) Prediction (interpolating in Ca-rich region and extrapolating in Al-rich region) of Young's modulus of CAS glasses using topology-informed ML approach over the entire CAS domain. The vertical dash line is the charge-compensate line, where $[Al_2O_3] = [CaO]$, to separate the two regions.

Table 3
Comparison between the interpolation and extrapolation ability with topology inputs for different machine learning algorithms used herein, namely, polynomial regression (PR), random forest (RF), and multilayer perceptron (MLP). The coefficient of determination (R^2) and root mean square error (RMSE) are used to describe the accuracy of predictions for training and test set.

Topology inputs	Interpolation				Extrapolation			
	R^2_train	R^2_test	RMSE_train	RMSE_test	R^2_train	R^2_test	RMSE_train	RMSE_test
PR	0.973	0.946	2.790	3.805	0.893	−89.64	2.238	145.2
RF	0.993	0.951	1.398	3.615	0.976	−0.806	1.252	20.494
MLP	0.962	0.945	3.330	3.819	0.832	0.903	4.026	4.655

MLP model successfully captures the difference in slope of the Young's modulus in the Ca-rich and Al-rich regions and exhibits a smaller confidence interval compared to the conventional MLP model in Fig. 3h The ternary plot in Fig. 4j also reveals the change in trends of Young's modulus from left to right, which closely aligns with our simulated results (see Fig. 1b).

Furthermore, we employ various conditions, as outlined in Table 4, to partition the dataset into training and test sets for extrapolation models. The performance comparison between simulated and predicted Young's modulus is shown in Fig 5. Each of these conditions results in a train-test split ratio of approximately 50–50. Subsequently, we train a topology-informed multi-layer perceptron (MLP) model to predict and extrapolate the Young's modulus of CAS glasses for each splitting method. The RMSE values for both the training and test sets are documented in Table 4. Our findings suggest that the topology-informed approach exhibits robust performance in extrapolation tasks, irrespective of the specific conditions utilized for dataset splitting. This observation underscores the versatility and effectiveness of our methodology in capturing the intrinsic relationships between the structural topology and the mechanical properties of CAS glasses, enabling accurate predictions even when confronted with novel compositional domains.

4. Discussion

We now focus on interpreting the model results by conducting a SHapley Additive exPlanation (SHAP) analysis and feature effect analysis with contour plots. SHAP analysis is a tool for explaining the output of ML models by quantifying the impact of each feature on the model's prediction [38]. Fig 5a shows a summary plot of the SHAP analysis for the topology-informed extrapolation MLP model. The vertical axis of a SHAP summary plot ranks the input features in terms of descending influence from top to bottom. A positive impact value indicates that a specific feature contributes positively to the model's prediction, and the magnitude of the SHAP value represents the strength of the features' influence on the model prediction. The value of each feature is color-coded. We observe that n_C/V has higher impact compared to BS/n_C, but both features exhibit positive-oriented impact on Young's modulus, i.e., higher values of features result in higher predicted Young's modulus. These observations come from the different physics information carried by the two topological inputs. n_C/V represents the volumetric density of the topological constraints, which has the majority contribution to the total connectivity of atomic glass structures [27]. BS/n_C describes the fraction of bond-stretching constraints in the total number of constraints, which presents the degree of connectivity of the glass [27]. This is consistent with our previous findings that both BS and BB constraints influence the stiffness of glass, but the BS constraints show a greater impact than the BB constraints [25].

In Figs. 6b and 6c, we show two contour plots to further evaluate the impact of features on the prediction of Young's modulus for both composition-informed interpolation MLP model (from Fig. 2) and topology-informed extrapolation MLP model (from Fig. 4h-j), respectively. For each plot, we generate 2-dimentional "fake" inputs, but covering the entire input domain, and output the corresponding predicted Young's modulus using the trained model. All predicted Young's modulus values are color-coded. The simulated Young's modulus values are plotted as square points for the training set and triangle points for the test set. We note that both models give accurate predictions for Young's modulus and well capture the composition-property and structure-property mappings. Most importantly, instead of capturing the non-linear composition-property mapping (i.e., the non-linear color pattern and contour lines in Fig. 6b) for the composition-informed approach, the topology-informed model only needs to find the linearized structure-property mapping (i.e., the linear color pattern and contour lines in Fig. 6c) with designed topological features. The intrinsic linearized mapping between the topological feature and Young's modulus of glass is the key to enable the extrapolation ability of the ML model in our work.

In summary, all the ML algorithms employed in this study demonstrate satisfactory performance in interpolation tasks, which aligns with our initial expectations. However, the intrinsic limitations inherent to the PR and RF algorithms hinder their ability to perform extrapolation predictions effectively. Our findings show that the MLP model, when combined with topology-informed features, proves to be the most suitable choice for extrapolation tasks. The success of our topology-informed approach in extrapolation is ascribed to the fact that topological features exhibit a high degree of linearity with the Young's modulus of glasses, as discussed previously. Nevertheless, implementing this approach requires expanding topological constraints theory or developing other readily calculable physical characteristics across a broad range of glass compositions. These features should aim to diminish the nonlinear relationship between a material's composition and its properties. This task remains both complex and essential, highlighting a critical area for future research and development in the field of materials science and machine learning integration.

Table 4
List of conditions to split the training and test set for extrapolation models, along with the corresponding root mean square error (RMSE) values for the training and test sets obtained after training a topology-informed MLP model under each condition. The table also lists the plots where the models' performance can be visualized.

	Conditions for extrapolation	RMSE_training (GPa)	RMSE_test (GPa)	Model performance visualization
1	[SiO_2] < 30 %	2.990	4.647	Fig 5a
2	[SiO_2] > 30 %	3.413	5.851	Fig 5b
3	[Al_2O_3] < 30 %	2.998	6.284	Fig 5c
4	[Al_2O_3] > 30 %	3.437	4.736	Fig 5d
5	[CaO] < 30 %	3.608	4.788	Fig 5e
6	[CaO] > 30 %	3.157	4.738	Fig 5f
7	([Al_2O_3] - [CaO]) < 0 %	3.879	5.213	Fig 5g
8	([Al_2O_3] - [CaO]) > 0 %	4.026	4.655	Fig 4h

5. Conclusion

In this study, we demonstrated that integrating topological knowledge derived from atomic structures substantially improves machine learning approaches for materials discovery, particularly enhancing extrapolation capabilities beyond conventional methods. By examining

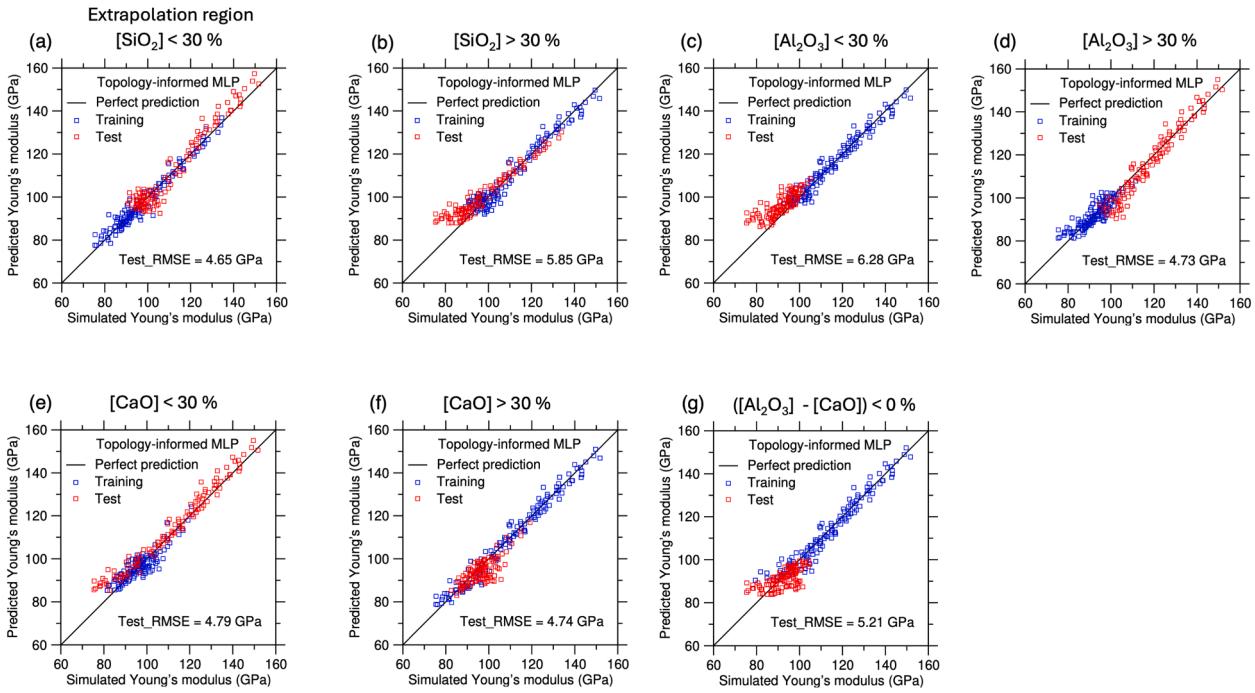

Fig. 5. Topology-informed MLP models extrapolate the Young's modulus of CAS glasses. The titles above each figure indicate the conditions for selecting the extrapolation dataset as the test set, while the remaining data are used as training set. The values of root mean square error on test set are shown in each figure.

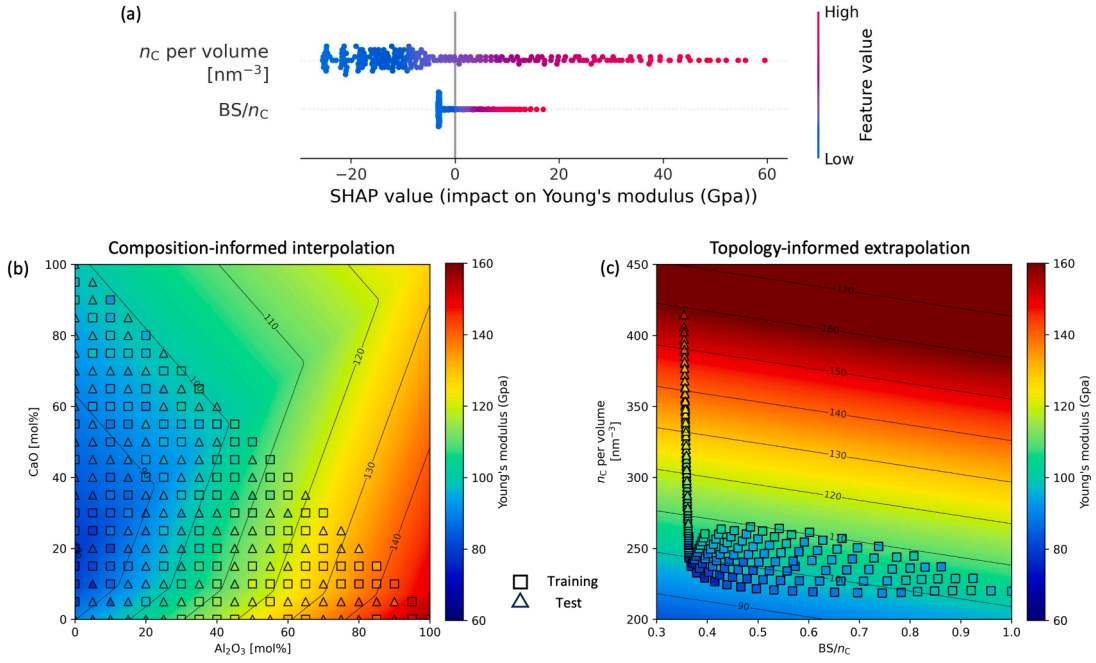

Fig. 6. Model interpretation. (a) SHAP summary plot showing the impact of each topological feature from the topology-informed model. (b) Contour plot showing the impact of features on the Young's modulus from the composition-informed MLP model with color-coded predictions. (c) Contour plot showing the impact of features on the Young's modulus from the topology-informed MLP model with color-coded predictions. Contour lines are drawn in both figures. Square and triangle points are color-coded and represent the simulated Young's modulus values used to train and test ML models, respectively.

Young's modulus prediction in CaO-Al_2O_3-SiO_2 glasses, we found that utilizing topological parameter converts a non-linear composition-property mapping into a more linear topology-property mapping. This transformation not only preserves prediction accuracy within the training domain but also significantly enhances the model's ability to extrapolate to unexplored compositional spaces. Our topology-informed machine learning approach provides an efficient pathway for discovering novel glass materials, effectively addressing a fundamental limitation of traditional data-driven methods that perform poorly in extrapolation tasks due to limited training datasets. This topology-informed strategy represents a promising direction for accelerating materials discovery in domains where experimental data remains sparse.

Data Availability

The data supporting the findings in this paper are available from the

corresponding author upon request.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used Claude 3.7 Sonnet in order to improve the readability and language of the manuscript. The authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Kai Yang: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation. **Yu Song:** Writing – review & editing, Visualization, Methodology. **Yuhai Li:** Methodology. **Morten M. Smedskjaer:** Writing – review & editing, Methodology. **Mathieu Bauchy:** Supervision, Methodology, Funding acquisition, Conceptualization. **Fabian Rosner:** Writing – review & editing, Visualization, Supervision, Investigation.

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 by the National Science Foundation (under Grants DMR-1928538)

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jnoncrysol.2025.123610.

References

[1] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, Accelerating materials property predictions using machine learning, Sci. Rep. 3 (2013) 2810, https://doi.org/10.1038/srep02810.
[2] J.E. Saal, A.O. Oliynyk, B. Meredig, Machine learning in Materials discovery: confirmed predictions and their underlying approaches, Annu. Rev. Mater. Res. 50 (2020) 49–69, https://doi.org/10.1146/annurev-matsci-090319-010954.
[3] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, NPJ. Comput. Mater. 3 (2017) 1–13, https://doi.org/10.1038/s41524-017-0056-5.
[4] L. Wondraczek, J.C. Mauro, J. Eckert, U. Kühn, J. Horbach, J. Deubener, T. Rouxel, Towards ultrastrong glasses, Adv. Mater. 23 (2011) 4578–4586, https://doi.org/10.1002/adma.201102795.
[5] T. Miyashita, T. Manabe, Infrared Optical fibers, IEEE Trans Microw. Theory Tech. 30 (1982) 1420–1438, https://doi.org/10.1109/TMTT.1982.1131275.
[6] T. Kokubo, Bioactive glass ceramics: properties and applications, Biomaterials 12 (1991) 155–163, https://doi.org/10.1016/0142-9612(91)90194-F.
[7] M.J. Plodinec, Borosilicate glasses for nuclear waste imobilisation, Glass Technol. 41 (2000) 186–192.
[8] H. Liu, Z. Fu, K. Yang, X. Xu, M. Bauchy, Machine learning for glass science and engineering: a review, J. Non Cryst. Solids 557 (2021) 119419, https://doi.org/10.1016/j.jnoncrysol.2019.04.039.
[9] K. Yang, X. Xu, B. Yang, B. Cook, H. Ramos, N.M.A. Krishnan, M.M. Smedskjaer, C. Hoover, M. Bauchy, Predicting the young's modulus of silicate glasses using high-throughput molecular dynamics simulations and machine learning, Sci. Rep. 9 (2019) 8739, https://doi.org/10.1038/s41598-019-45344-3.
[10] Y.-J. Hu, G. Zhao, M. Zhang, B. Bin, T. Del Rose, Q. Zhao, Q. Zu, Y. Chen, X. Sun, M. de Jong, L. Qi, Predicting densities and elastic moduli of SiO_2-based glasses by machine learning, NPJ. Comput. Mater. 6 (2020) 1–13, https://doi.org/10.1038/s41524-020-0291-z.
[11] R. Ravinder, K.H. Sridhara, S. Bishnoi, H.S. Grover, M. Bauchy, H. Kodamana Jayadeva, N.M.A. Krishnan, Deep learning aided rational design of oxide glasses, Mater. Horiz. 7 (2020) 1819–1827, https://doi.org/10.1039/D0MH00162G.
[12] D.R. Cassar, A.C.P.L.F. de Carvalho, E.D. Zanotto, Predicting glass transition temperatures using neural networks, Acta Mater. 159 (2018) 249–256, https://doi.org/10.1016/j.actamat.2018.08.022.
[13] N.M. Anoop Krishnan, S. Mangalathu, M.M. Smedskjaer, A. Tandia, H. Burton, M. Bauchy, Predicting the dissolution kinetics of silicate glasses using machine learning, J. Non Cryst. Solids 487 (2018) 37–45, https://doi.org/10.1016/j.jnoncrysol.2018.02.023.
[14] D.R. Cassar, ViscNet: neural network for predicting the fragility index and the temperature-dependency of viscosity, Acta Mater. 206 (2021) 116602, https://doi.org/10.1016/j.actamat.2020.116602.
[15] K. Gong, E. Olivetti, Data-driven prediction of room-temperature density for multicomponent silicate-based glasses, J. Am. Ceramic Soc. 106 (2023) 4142–4162, https://doi.org/10.1111/jace.19072.
[16] B. Meredig, E. Antono, C. Church, M. Hutchinson, J. Ling, S. Paradiso, B. Blaiszik, I. Foster, B. Gibbons, J. Hattrick-Simpers, A. Mehta, L. Ward, Can machine learning identify the next high-temperature superconductor? Examining extrapolation performance for materials discovery, Mol. Syst. Des. Eng. 3 (2018) 819–825, https://doi.org/10.1039/C8ME00012C.
[17] L.M. Ghiringhelli, J. Vybiral, S.V. Levchenko, C. Draxl, M. Scheffler, Big data of materials science: critical role of the descriptor, Phys. Rev. Lett. 114 (2015) 105503, https://doi.org/10.1103/PhysRevLett.114.105503.
[18] L.M. Ghiringhelli, J. Vybiral, E. Ahmetcik, R. Ouyang, S.V. Levchenko, C. Draxl, M. Scheffler, Learning physical descriptors for materials science by compressed sensing, New. J. Phys. 19 (2017) 023017, https://doi.org/10.1088/1367-2630/aa57bf.
[19] R. Tibshirani, Regression shrinkage and selection via the lasso, J. R. Stat. Soc. Ser. B (Methodological) 58 (1996) 267–288.
[20] M. Verleysen, D. François, The curse of dimensionality in data mining and time series prediction, in: J. Cabestany, A. Prieto, F. Sandoval (Eds.), Computational Intelligence and Bioinspired Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005: pp. 758–770. https://doi.org/10.1007/11494669_93.
[21] J.C. Phillips, Topology of covalent non-crystalline solids I: short-range order in chalcogenide alloys, J. Non Cryst. Solids 34 (1979) 153–181, https://doi.org/10.1016/0022-3093(79)90033-4.
[22] J.C. Phillips, Topology of covalent non-crystalline solids II: medium-range order in chalcogenide alloys and A-Si(Ge), J. Non Cryst. Solids 43 (1981) 37–77, https://doi.org/10.1016/0022-3093(81)90172-1.
[23] M. Bauchy, Deciphering the atomic genome of glasses by topological constraint theory and molecular dynamics: a review, Comput. Mater. Sci. 159 (2019) 95–102, https://doi.org/10.1016/j.commatsci.2018.12.004.
[24] M.M. Smedskjaer, J.C. Mauro, Y. Yue, Prediction of glass hardness using temperature-dependent constraint theory, Phys. Rev. Lett. 105 (2010) 115503, https://doi.org/10.1103/PhysRevLett.105.115503.
[25] K. Yang, B. Yang, X. Xu, C. Hoover, M.M. Smedskjaer, M. Bauchy, Prediction of the Young's modulus of silicate glasses by topological constraint theory, J. Non Cryst. Solids 514 (2019) 15–19, https://doi.org/10.1016/j.jnoncrysol.2019.03.033.
[26] J.C. Mauro, Topological constraint theory of glass, Am. Ceramic Soc. Bull. 90 (2011) 31.
[27] Y. Hu, H. Liu, K. Yang, Q. Zhou, C.G. Hoover, N.M.A. Krishnan, M.M. Smedskjaer, M. Micoulaut, L. Guo, M. Bauchy, Topological constraint theory of glass: counting constraints by molecular dynamics simulations, in: Atomistic Simulations of Glasses, John Wiley & Sons, Ltd, 2022: pp. 123–148. https://doi.org/10.1002/9781118939079.ch5.
[28] H. Liu, T. Zhang, N.M. Anoop Krishnan, M.M. Smedskjaer, J.V. Ryan, S. Gin, M. Bauchy, Predicting the dissolution kinetics of silicate glasses by topology-informed machine learning, Npj. Mater. Degrad. 3 (2019) 32, https://doi.org/10.1038/s41529-019-0094-1.
[29] J.C. Mauro, Decoding the glass genome, Curr. Opin. Solid State Mater. Sci. 22 (2018) 58–64, https://doi.org/10.1016/j.cossms.2017.09.001.
[30] M.M. Smedskjaer, M. Bauchy, Sub-critical crack growth in silicate glasses: role of network topology, Appl. Phys. Lett. 107 (2015) 141901, https://doi.org/10.1063/1.4932377.
[31] T. Oey, E. Callagon, G. Falzone, K. Yang, A. Wada, M. Bauchy, J. Bullard, G. Sant, Topological controls on aluminosilicate glass dissolution: complexities induced in hyperalkaline aqueous environments, J. Am. Ceram. Soc. 103 (2020) 6198–6207.
[32] Y. Song, K. Yang, J. Chen, K. Wang, G. Sant, M. Bauchy, Machine learning enables rapid screening of reactive fly ashes based on their network topology, ACS Sustain. Chem. Eng 9 (2021) 2639–2650, https://doi.org/10.1021/acssuschemeng.0c06978.
[33] K. Yang, Y. Hu, Z. Li, N.M.A. Krishnan, M.M. Smedskjaer, C.G. Hoover, J.C. Mauro, G. Sant, M. Bauchy, Analytical model of the network topology and rigidity of calcium aluminosilicate glasses, J. Am. Ceramic Soc. 104 (2021) 3947–3962, https://doi.org/10.1111/jace.17781.
[34] L. Chen, H. Tran, R. Batra, C. Kim, R. Ramprasad, Machine learning models for the lattice thermal conductivity prediction of inorganic materials, Comput. Mater. Sci. 170 (2019) 109155, https://doi.org/10.1016/j.commatsci.2019.109155.
[35] Y.T. Sun, H.Y. Bai, M.Z. Li, W.H. Wang, Machine learning approach for prediction and understanding of glass-forming ability, J. Phys. Chem. Lett. 8 (2017) 3434–3439, https://doi.org/10.1021/acs.jpclett.7b01046.
[36] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, https://doi.org/10.1038/nature14539.
[37] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[38] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, in: Advances in Neural Information Processing Systems, Curran Associates, Inc., 2017. https://proceedings.neurips.cc/paper, 2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (accessed April 16, 2024).

[39] A. Roth, L. Shapley, The Shapley value, Cambridge University Press, 1988. https://books.google.com/books/about. The_Shapley_value.html?id=JK7MKu2A9cIC (accessed April 16, 2024).
[40] M. Kazembeyki, K. Yang, J.C. Mauro, M.M. Smedskjaer, M. Bauchy, C.G. Hoover, Decoupling of indentation modulus and hardness in silicate glasses: evidence of a shear- to densification-dominated transition, J. Non Cryst. Solids 553 (2021) 120518, https://doi.org/10.1016/j.jnoncrysol.2020.120518.
[41] M. Bauchy, Structural, vibrational, and elastic properties of a calcium aluminosilicate glass from molecular dynamics simulations: the role of the potential, J. Chem. Phys. 141 (2014) 024507, https://doi.org/10.1063/1.4886421.
[42] S. Takahashi, D.R. Neuville, H. Takebe, Thermal properties, density and structure of percalcic and peraluminus CaO–Al2O3–SiO2 glasses, J. Non Cryst. Solids 411 (2015) 5–12, https://doi.org/10.1016/j.jnoncrysol.2014.12.019.
[43] M.J. Toplis, D.B. Dingwell, T. Lenci, Peraluminous viscosity maxima in Na2O Al2O3 SiO2 liquids: the role of triclusters in tectosilicate melts, Geochim. Cosmochim. Acta 61 (1997) 2605–2612, https://doi.org/10.1016/S0016-7037(97)00126-9.
[44] Y. Xiang, J. Du, M.M. Smedskjaer, J.C. Mauro, Structure and properties of sodium aluminosilicate glasses from molecular dynamics simulations, J. Chem. Phys. 139 (2013), https://doi.org/10.1063/1.4816378, 044507.
[45] P.K. Gupta, J.C. Mauro, Composition dependence of glass transition temperature and fragility. I. A topological model incorporating temperature-dependent constraints, J. Chem. Phys. 130 (2009) 094503, https://doi.org/10.1063/1.3077168.
[46] J.C. Mauro, P.K. Gupta, R.J. Loucks, Composition dependence of glass transition temperature and fragility. II. A topological model of alkali borate liquids, J. Chem. Phys. 130 (2009) 234503, https://doi.org/10.1063/1.3152432.