Journal of Non-Crystalline Solids: X 4 (2019) 100036

Contents lists available at ScienceDirect

Journal of Non-Crystalline Solids: X

journal homepage: www.journals.elsevier.com/journal-of-non-crystalline-solids-x

Machine learning for glass science and engineering: A review

Han Liu[a], Zipeng Fu[a,b], Kai Yang[a], Xinyi Xu[a,c], Mathieu Bauchy[a,*]

[a] Physics of AmoRphous and Inorganic Solids Laboratory (PARISlab), Department of Civil and Environmental Engineering, University of California, Los Angeles, CA 90095, USA
[b] Department of Computer Science, University of California, Los Angeles, CA 90095, USA
[c] Department of Mathematics, University of California, Los Angeles, CA 90095, USA

ARTICLE INFO

Keywords:
Composition-property relationship
Structural signature
Molecular dynamics simulation
Artificial neuron network
Bayesian optimization

ABSTRACT

The design of new glasses is often plagued by poorly efficient Edisonian "trial-and-error" discovery approaches. As an alternative route, the Materials Genome Initiative has largely popularized new approaches relying on artificial intelligence and machine learning for accelerating the discovery and optimization of novel, advanced materials. Here, we review some recent progress in adopting machine learning to accelerate the design of new glasses with tailored properties.

1. Introduction

1.1. Challenges in the development of new glasses

Developing novel glasses with new, improved properties and functionalities is key to address some of the Grand Challenges facing our society [1,2]. Although the process of designing a new material is always a difficult task, the design of novel glasses comes with some unique challenges. First, virtually all the elements of the periodic table can be turned into a glass if quenched fast enough [3]. Second, unlike crystals, glasses are intrinsically out-of-equilibrium and, hence, can exhibit a continuous range in their stoichiometry (within the glass-forming ability domain) [4]. For both of these reasons, the compositional envelope that is accessible to glass is limitless and, clearly, only an infinitesimal fraction of these compositions have been explored thus far [3]. Although the vast compositional envelop accessible to glass opens endless possibilities for the discovery of new glasses with unusual properties, efficiently exploring such a high-dimension space is notoriously challenging and traditional discovery methods based on trial-and-error Edisonian approaches are highly inefficient [5]. Although "intuition" can partially overcome these challenges, it is unlikely to yield a leapfrog in glass properties and functionalities.

As a first option, physics-based modeling can greatly facilitate the design of new glasses by predicting a range of optimal promising compositions to focus on [6]. For instance, topological constraint theory has led to the development of several analytical models predicting glass properties as a function of their compositions (e.g., glass transition temperature, hardness, stiffness, etc.) [7–12]. However, the complex, disordered nature of glasses renders challenging the development of accurate and transferable physics-based models for certain properties (e.g., liquidus temperature, fracture toughness, dissolution kinetics, etc.) [6]. Alternatively, "brute-force" atomistic modeling techniques (e.g., molecular dynamics) can be used to accurately compute glass properties and partially replace more costly experiments (see also Section 3.5) [13,14]. However, such techniques come with their own challenges (e.g., limited timescale, small number of atoms, fast cooling rate, large computing cost, etc.), which prevents a systematic exploration of all the possible glasses [15–17].

1.2. When machine learning meets glass science

As an alternative route to physics-based modeling, artificial intelligence and machine learning offer a promising path to leverage existing datasets and infer data-driven models that, in turn, can be used to accelerate the discovery of novel glasses [11,18]. As a notable success, machine learning modeling techniques have been used to accelerate the design of Corning® Gorilla® glasses [18]. Over the past decade, thanks to the rapid increase in available computing power, artificial intelligence and machine learning have revolutionized various aspects of our lives [19,20], including for image recognition [21], Internet data mining [22], or self-driving cars [23].

In details, machine learning can "learn from example" by analyzing existing datasets and identifying patterns in data that are invisible to human eyes [24]. Fig. 1 shows a typical application of machine learning to glass design. First, some data are generated (by experiments, simulations, or mining from existing databases) to build a database of

* Corresponding author.
E-mail address: bauchy@ucla.edu (M. Bauchy).

https://doi.org/10.1016/j.nocx.2019.100036

Available online 22 August 2019

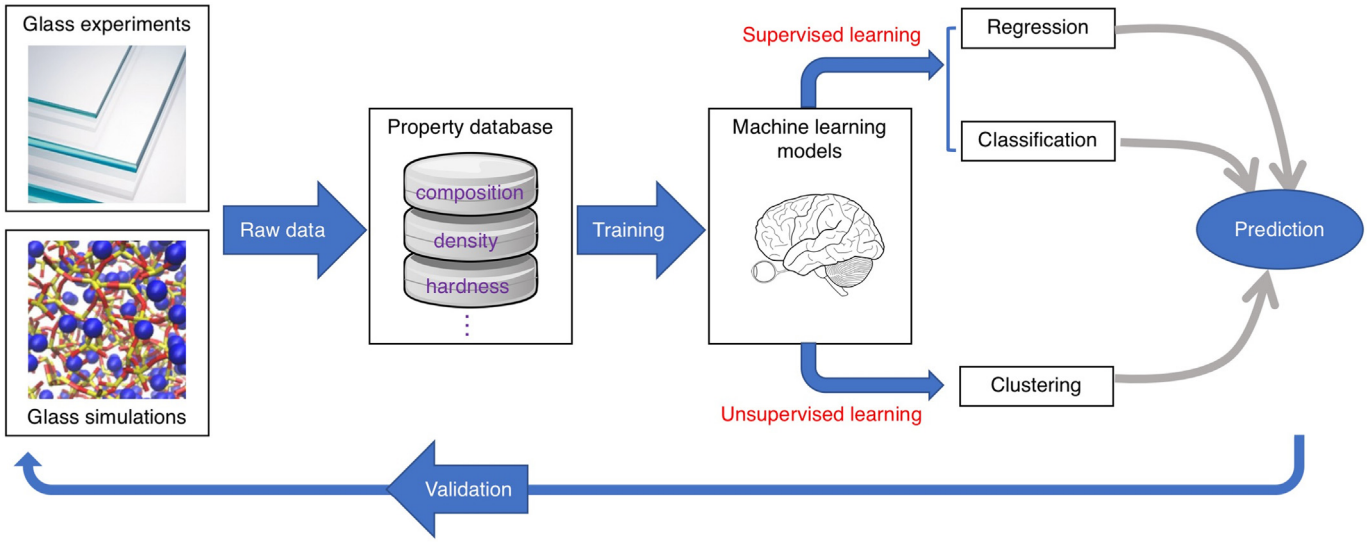

Fig. 1. Illustration of a typical application of machine learning to facilitate glass design.

properties. Such databases can comprise, as an example, the glass composition, synthesis procedure, as well as select properties. Machine learning is then used to infer some patterns within the dataset and establish a predictive model [24].

Machine learning algorithms can accomplish two types of tasks, namely, supervised and unsupervised. In the case of supervised machine learning, the dataset comprises a series of inputs (e.g., glass composition) and outputs (e.g., density, hardness, etc.). Supervised machine learning can then learn from these existing examples and infer the relationship between inputs and outputs [25]. Supervised machine learning comprises (i) regression algorithms [26], which can be to predict the output as a function of the inputs (e.g., composition-property predictive models) and (ii) classification algorithms [27], which can be used to label glasses within different categories. In contrast, in the case of unsupervised machine learning, the dataset is not labeled (i.e., no output information is known) [28]. Unsupervised machine learning can, for instance, be used to identify some clusters within existing data, that is, to identify some families of data points that share similar characteristics [29]. More details about these machine learning methods are presented in Section 2.

1.3. Challenges and limitations of machine learning for glass science

Although machine learning offers a unique, largely untapped opportunity to accelerate the discovery of novel glasses with exotic functionalities, it faces several challenges. First, the use of machine learning requires as a prerequisite the existence of data that are (i) available (i.e., public and easily accessible), (ii) complete, (iii) consistent (e.g., obtained from a single operation), (iv) accurate (i.e., with low error bars [30]), and (v) numerous [31]. For instance, although some glass property databases are available [32], inconsistencies between data generated by different groups render challenging the meaningful application of machine learning approaches. In addition, since they are usually only driven by data and do not embed any physics- or chemistry-based knowledge, machine learning models can sometimes violate the laws of physics or chemistry [33,34]. For these reasons, conventional machine learning techniques are usually good at "interpolating" data, but have thus far a limited potential for "extrapolating" predictions far from their initial training set [34,35], which usually prevents the efficient exploration of new unknown compositional domains (see Section 3.2 on how "physics-informed machine learning" can offer improved extrapolations). Finally, machine learning models often offer poor interpretability, that is, they act as black boxes and do not offer clear physical insights [36–38]. Here, we review some recent progress aiming to address and mitigate these challenges.

This review is organized as follows. First, Section 2 presents an overview of available machine learning techniques. Section 3 then reviews the state-of-the-art in the application of machine learning to glass science and engineering. Finally, Section 4 offers some conclusions and future directions.

2. Overview of machine learning techniques for glass science

2.1. Regression techniques

2.1.1. Parametric and nonparametric regression

Regression consists of fitting known data points to establish a functional relationship between the inputs and output [26]. As illustrated in Fig. 2a, regression models are able to interpolate known points by learning from an existing dataset. Generally, regression methods can be categorized into (i) parametric regression, which yields an analytical formula expressing the output in terms of the input variables [26] (e.g., linear [39], polynomial [40], or nonlinear functions [41]) and (ii) nonparametric regression, which defines a kernel function to calculate the output at a given input position based on the correlation between this input position and its surrounding known points [42].

Nonparametric regression comprises, for instance, the K-nearest-neighbor (KNN) [43] and Gaussian process regression (GPR) methods [44]. The basic idea of the KNN method is to predict the value of the output for a given input position by using the average value of the K nearest known points at the vicinity of the input position. On the other hand, the GPR method predicts a Gaussian-type probability distribution of the output for each input position based on the multivariate normal correlation between this input position and all the other known points [45]—wherein the degree of correlation decreases as a function of the distance between these points [44]. As a major advantage, the GPR method is able to provide the uncertainty of the predicted output values, which is key to assess the reliability of the predictions [46].

In contrast to nonparametric regression, parametric regression relies on an explicit analytical formula relating the inputs to the output—wherein the parameters of the formula are adjusted to fit the known points by establishing and minimizing a cost function [26]. It is worth pointing out that more complex machine learning algorithms (described in Section 2.3) can be used for classification and regression. For instance, artificial neuron network (ANN) [19,47], support vector

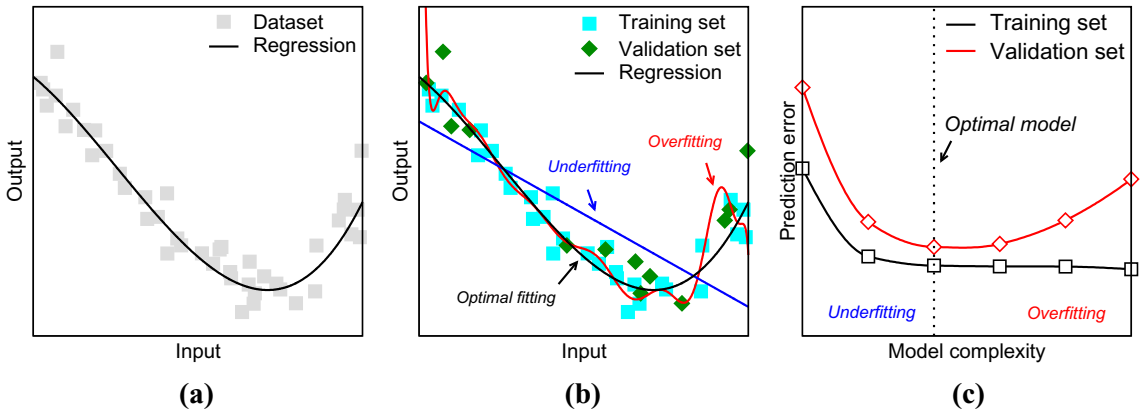

Fig. 2. Illustration of regression machine learning techniques. (a) Example of regression (black line) applied on an existing dataset (grey points). For illustration purposes, a polynomial regression model (with a polynomial degree $p = 3$) is adopted herein. (b) Illustration of underfitting (blue line, $p = 1$) and overfitting (red line, $p = 15$) on the same dataset. The dataset is divided into a (i) training set (cyan points), which is used to train the model, and (ii) validation set (green points), which is used to estimate how well the model can predict data that are kept invisible during its training. (c) Error in the prediction of the training (black line) and validation (red line) sets as a function of the model complexity (i.e., p in the polynomial model herein). The optimal model presents the lowest validation set prediction error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

machine (SVM) [48], random forest [49], or gradient boosting [50] essentially rely on complex nonlinear parametric formulas and, hence, can be classified as parametric regression techniques, except in the case of kernel-based functions [51]. These types of models often show a very good ability to interpolate data [52], but usually present low interpretability due to the complex format of the parametric formula [47] and limited extrapolation abilities [35].

2.1.2. Optimization of model complexity

The development of supervised learning models usually comprises two stages, viz., (i) the learning/fitting (i.e., training and validation) stage and (ii) the prediction (i.e., test) stage. During the fitting/learning stage, it is key to properly adjust the complexity of the model (e.g., the maximum degree in polynomial regression) to offer reliable predictions [53,54]. This process is described in the following.

Underfitting and overfitting: In the case of underfitting (i.e., low complexity), the model is too simple to properly capture the functional relationship between the inputs and output. In contrast, in the case of overfitting (i.e., high complexity), the model keeps the memory of the "noise" of the dataset [55]. In general, the model complexity can be captured by the number of non-zero fitting parameters, number of inputs, and number of high-order terms in a model [20,24]. Fig. 2b illustrates the manifestations of underfitting and overfitting by fitting a set of data (i.e., training set, see below) when some polynomial models with varying maximum polynomial degrees p. Clearly, in this case, a linear model with $p = 1$ does not properly capture the non-linear relationship between inputs and output. In contrast, a polynomial model with $p = 15$ is able to capture the noise of the training set, which, in turn, yields a poor predicting for unknown data points (i.e. validation set, see below). In between these two regimes, a polynomial regression model with $p = 3$ is able to properly capture the trend of the data while filtering out the noise of the dataset.

Training, validation, and test sets: To limit the risk of overfitting and assess the accuracy of the model, the dataset is usually divided into the training, validation, and test sets [20,24]. The training set is first used to train the model, that is, to adjust the model parameters in order to fit some existing data points. At this stage, the training and test sets are kept fully invisible to the model. Afterward, the validation set is used to adjust the complexity of the model. Indeed, as illustrated in Fig. 2c, higher model complexity (i.e., higher p herein) usually yields an improved interpolation of the training set, but eventually results in a lower ability to predict the training set as the model starts to remember the noise of the training set. Overall, the optimal degree of complexity manifests itself by a minimum prediction error for the validation set [55]. Finally, once the optimal degree of complexity is fixed, the test set is used to assess the accuracy of the model by comparing its predictions to a fraction of the dataset that is kept unknown to the model.

K-fold cross-validation: In many realistic cases, the limited number of data present in datasets makes it undesirable to keep a large fraction of the data fully unknown to the model as a validation—since a large number of data points is key to ensure the proper training of the model. This challenge can be overcome by using the K-fold cross-validation technique [24,56]. This technique divides the initial training set into K folds, trains the model based on K – 1 of the folds, and uses the remaining fold for validation. This procedure is then repeated K times until each of the folds has been used as a validation set. The accuracy of the model is then determined by averaging the accuracy of the prediction over all the K validation folds.

Regularization methods: An alternative route to decrease the model complexity consists in filtering out non-important terms from the model, which can be accomplished by regularization methods [57], e.g., LASSO [58], Ridge [59], or Elastic Net [57]. The main idea of regularization methods is to formulate and minimize a cost function that comprises (i) how well the model can predict known data as well as (ii) an additional term that attributes a penalty to complex models. As such, the minimization of the cost function forces non-important terms (i.e., which do not significantly contribute to increasing the accuracy of the model) to become zero. The degree of complexity of the model can be tuned by adjusting the weight attributed to the penalty term until the model offers an optimal prediction of the validation set [24,57].

2.2. Classification techniques

Classification can be viewed as a special case of regression [27]. In contrast to the case of regression—wherein the output is a continuous value—classification considers problems where the output is discrete, wherein each state corresponds to the labels to distinct categories. For instance, in the case of a binary classification problem, the data points belong to two classes (Class A and B), which can be represented by an output value equal to $+1$ or -1 for Class A and B, respectively. The goal of classification models is to predict the class of unknown data (e.g., "glass is transparent or not transparent") as a function of the inputs (e.g., glass composition). This can be accomplished by identifying the optimal hyperplane within the inputs space that best divides the different classes (see Fig. 3) [20,24,27].

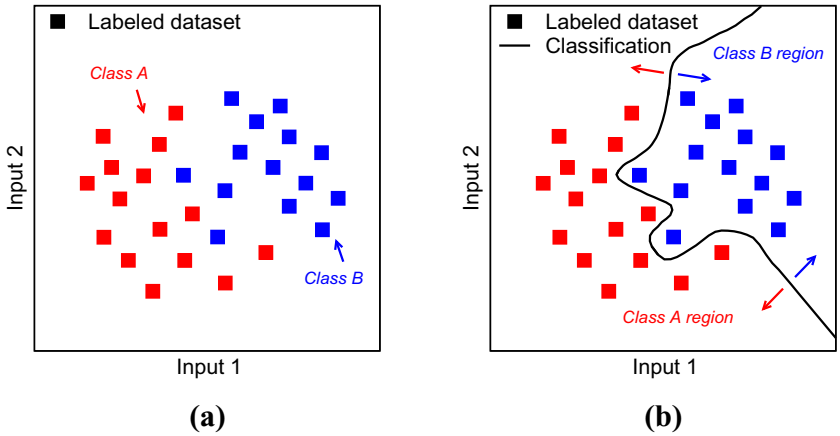

Fig. 3. Illustration of classification machine learning techniques. (a) Example of a dataset comprising two inputs (i.e., two-dimensional input space). The data points are labeled as belonging to either Class A (red points) or Class B (blue points). (b) Example of classification in the two-dimensional space. For illustration purposes, a support vector machine (SVM) model is adopted, which yields a hyperplane boundary (black line) that divides the two-dimensional space into two different class regions, i.e., Class A (left) and Class B (right). Note that a hyperplane has a dimensionality that is 1 degree lower than that of the input space and, as such, takes the form a line in a two-dimensional input space. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2.3. Examples of supervised machine learning algorithms

Artificial neural network (ANN): ANN algorithms, e.g., multilayer perceptron [60] or convolutional neural network (CNN) [21], rely on a multilayer structure comprising (i) an input monolayer, (ii) some hidden multilayer, and (iii) an output monolayer (see Fig. 5a). Each layer is made up of several neurons that are connected to each other to mimic the human neuron network system. Each neuron consists of a non-linear transformation operator (e.g., a sigmoid function) that relates the signal coming from the neurons from the previous layer to a response signal that is transmitted to the neurons of the subsequent layer. ANN can be viewed as a complex, non-linear functional mapping the relationship between the inputs and output(s) [25].

Support vector machine (SVM): SVM algorithms, which include both linear SVM [48] and kernel SVM [51], rely on a functional formula that represents the hyperplane that divides data into different classes in classification problems (see Section 2.2). On the one hand, linear SVM uses linear functions to express a set of linear hyperplanes to divide the input space into different class regions. The coefficients of the linear functions are determined by maximizing the separation/margin of the nearest known points on both sides of the hyperplane [48]. On the other hand, kernel SVM uses a kernel function that describes the correlation between an input position and the known points from the training set (i.e., for which the class is known). This yields a set of non-linear hyperplanes that can be used for classification. The parameters in the kernel function are also determined so as to maximize the margin [51].

Decision tree: Tree-based models, e.g., random forest [49], are based on an ensemble of several parallel tree paths made of sequentially splitting nodes. Each node represents a judge condition that guides the choice of the next node derived from it. The judge condition at each node, which can be expressed as a split of a target input range, is optimized based on the training set. Each parallel tree path gives its own predicted output and the final output value is determined from the overall votes from the outputs of all the tree paths. The tree size (i.e., the number of nodes) depends on the size of the dataset (in terms of the number of data points or the dimensionality of the input space). This parameter can be optimized by minimizing the prediction error of the validation set (see Fig. 2c), that is, to avoid both underfitting and overfitting [49].

Boosting method: Boosting models, e.g., AdaBoost [61] or gradient boosting [50], are based on an ensemble of sequentially-added weak learners/classifiers (e.g., decision tree, SVM, or other classifiers). In this case, the predicted output is given by a weighted average of the outputs yielded by all the weak learners/classifiers. Each weak learner is added in sequence and is mainly trained by the remaining training samples that are not well predicted from the weighted average of all the outputs of the previous weak learners. The weight coefficient attached to each weak learner, which represents its contribution to the final prediction, is determined from the updated prediction error of the assembled model after adding this weak learner [50].

2.4. Unsupervised machine learning—Clustering

Rather than learning by example (i.e., supervised machine learning), unsupervised machine learning aims to decipher some intrinsic characteristics of the input dataset itself. A typical example of unsupervised machine learning is the detection of clusters within data—wherein a cluster is a group of data that present similar characteristics [29]. In this case, no examples of previously identified

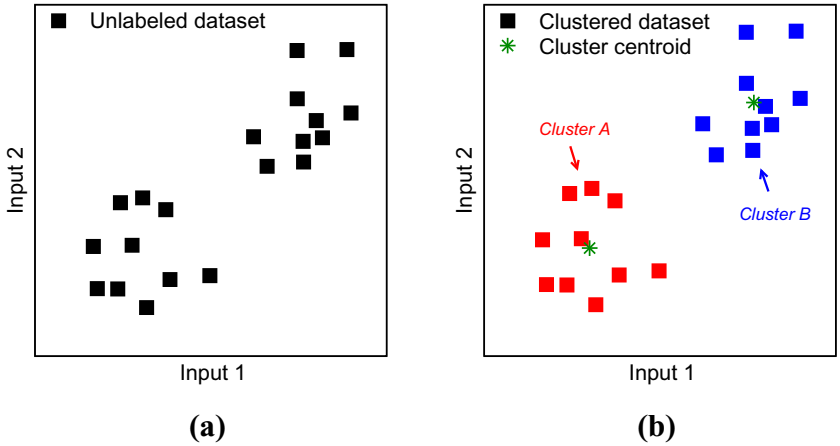

Fig. 4. Illustration of clustering machine learning techniques. (a) Example of a dataset comprising two inputs (i.e., two-dimensional input space). The data points (black points) are distributed within the inputs space in a non-homogeneous fashion. (b) Outcome of the clustering analysis, wherein the data points are labeled as belonging to clusters A (red points) or B (blue points). The centroid (green star) of each cluster is also shown. The clustering method used herein is the K-mean algorithm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

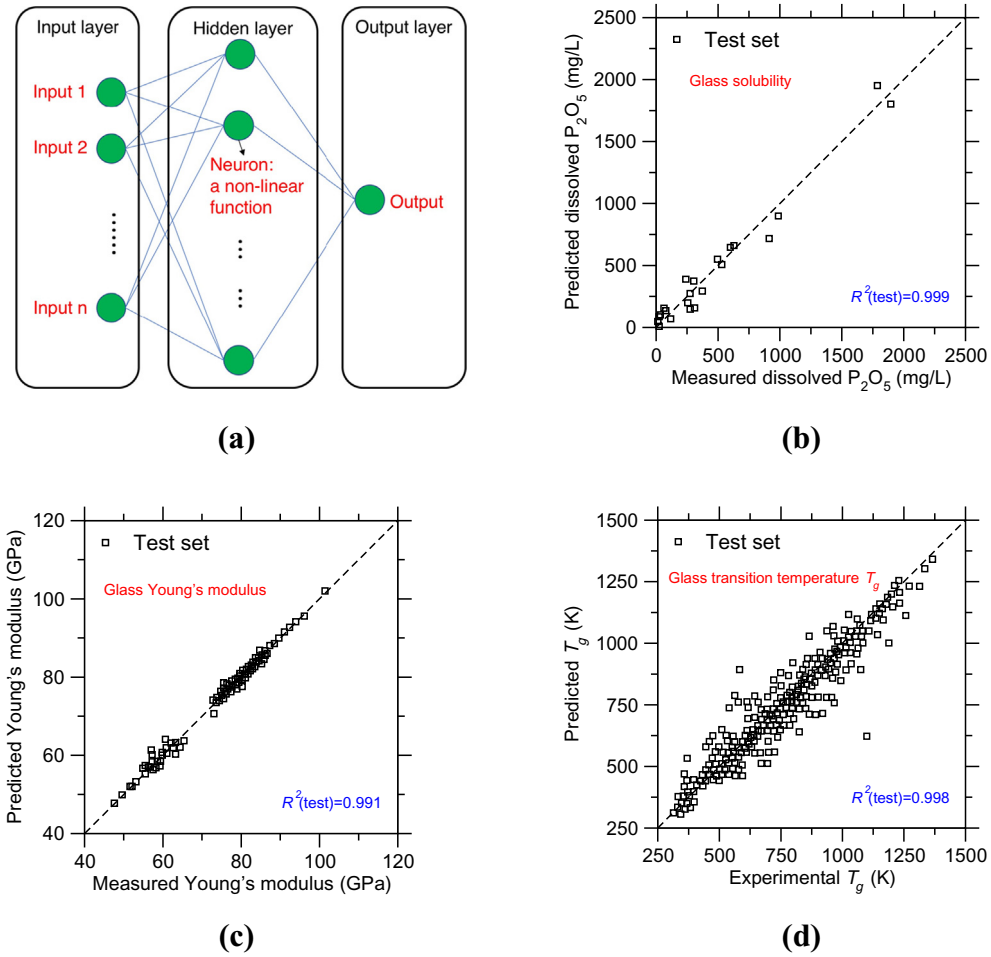

Fig. 5. (a) Illustration of an artificial neural network model, which comprises an input layer, hidden layer, and output layer. Here, the input variables refer to the glass composition. Comparison between predicted (i.e., the output of the model) and measured glass properties for (b) glass solubility [75], (c) Young's modulus [11], and (d) glass transition temperature [37]. The correlation coefficient R^2 is indicated as a measure of the model accuracy.

clusters are needed to train the model—and relevant clusters are identified based on the analysis of the distances between the data points within the inputs space. Fig. 4 shows an example of clustering analysis in a two-dimensional input space. In this case, based on the spatial distribution of the data, two clusters are detected (see Fig. 4b) [62].

The K-mean algorithm (and its derivations) is one of the most widely used clustering algorithm [63,64]. The basic idea of this algorithm is to first randomly place K clusters centroids within input space. At the first iteration, all the data points are labeled with a cluster ID (ranging from 1 to K) based on the ID of the cluster centroid they are the closest to. The position of each cluster centroid is then updated based on the average position of the labeled data points belonging to that cluster and all the data points are relabeled accordingly. This procedure is then iteratively repeated until the positions of each centroid converges and does not move any longer [65]. Note that, in the K-mean algorithm, the number of clusters K is fixed and is a prerequisite input of the model. However, several methods have been developed to determine the optimal number of parameters K [66], such as the Elbow method [67]—wherein the idea is to select an optimal value for K for which any further addition of centroids does not significantly reduce the cost function to be minimized (e.g., the square sum of the distances between each data point and its associated cluster centroid [67]). A common issue of the K-mean algorithm is that the algorithm remains stuck in a local minimum of the cost function during the optimization and does not converge to the global minimum [68]. This limitation can be partially overcome by repeating the clustering analysis several times while considering different random initial positions for the cluster centroids [68].

2.5. Feature engineering and dimensionality reduction

In both supervised and unsupervised learning, feature engineering is key to identify relevant inputs describing each data point (e.g., glass composition, synthesis method, annealing temperature, etc.) [24]. Each input variable is called a feature. To select the proper independent input variables, one must identify the system features that present the largest influence on the target output. This step is called feature engineering, which can be based on some physical knowledge of the problem or a statistical analysis of the correlation between inputs and output [69]. In practice, the relevant inputs can be identified based on some feature evaluation methods (for instance, by calculating the covariance matrix) [69]. However, in some cases, there are tens to hundreds of possible input variables that can be defined for a given data point—and such a high dimensionality of the input space would significantly reduce the computational efficiency of machine learning models [70]. To overcome the "curse of dimensionality" [24], some dimensionality reduction methods can be used to reduce the dimensionality of the inputs space, that is, to reduce the number of inputs considered during the training of the model. Such techniques include principal component analysis (PCA) [71], non-negative matrix factorization (NMF) [72], and linear discriminant analysis (LDA) [73]. Briefly, the main idea behind these methods is to use some linear/nonlinear combinations of the different available inputs to construct informative new inputs and replace some of the original inputs

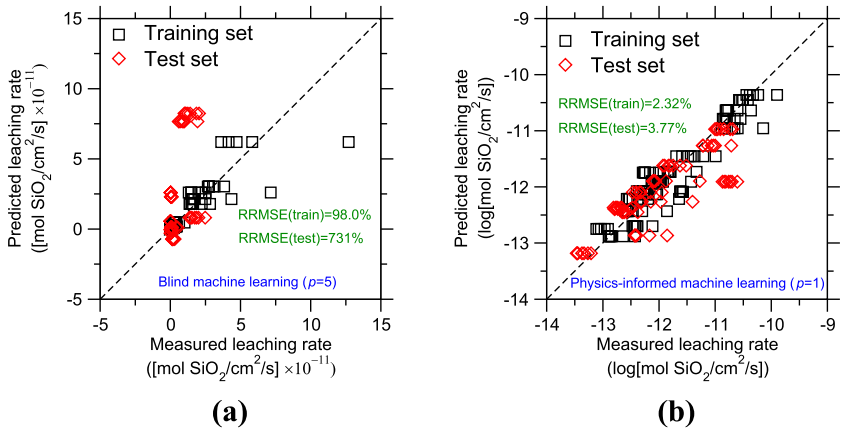

Fig. 6. Comparison between predicted and measured glass dissolution rates values, as offered by (a) "blind machine learning" and (b) "physics-informed machine learning" using polynomial regression models for the training and test sets [34]. Note that panel (b) presents the logarithm of the dissolution rate values.

[20,24,74]. As such, by combining several inputs into some single metrics, such techniques can be used to reduce the dimensionality of the model and, hence, enhance the computational efficiency of machine learning. It is worth pointing out that the minimum number of data points that is needed to train a machine learning model increases with increasing dimensionality—but also depends on the type of machine learning methods that is used, as well as the nature of the predicted property. Empirically, at least 3-to-5 data points per input dimension are required to meaningfully train a machine learning model.

3. Application of machine learning to glass science and engineering

3.1. Conventional composition-property regression models

Most applications of machine learning for glass science have focused on the development of composition-property regression models. To this end, pioneering works have focused on the use of the artificial neural network method (see Section 2.3 and Fig. 5a) [11,36,37,75]. To the best of our knowledge, the first use of machine learning in the context of glass science was conducted by Brauer et al. and aimed to predict the solubility of P_2O_5–CaO–MgO–Na_2O–TiO_2 glass as a function of composition [75]. Fig. 5b shows a comparison between the predicted and measured solubility. Overall, the predictions match well with experiments and the trained model yields a correlation coefficient R^2 for the test set that approaches 0.999 [76]. Following the pioneering work, various studies have focused on applying the artificial neural network method to predict the properties of glasses as a function of their composition [11,18,36,37,77]. As an example, Fig. 5c shows a comparison between predicted and measured values of the Young's modulus of a wide range of silicate glasses from a study conducted by Mauro et al.—wherein the model yields a correlation coefficient $R^2 = 0.991$ for the test set [11]. Finally, Fig. 5d shows the outcome of a recent work from Casser et al. wherein artificial neural network was used to predict the glass transition temperature (T_g) as a function of glass composition (with $R^2 = 0.998$ for the test set) [37]. This work exemplifies the ability of artificial neural network to handle complex datasets—since the glass transition temperature presents several definitions and is not consistently measured among different research groups [37]. This demonstrates the ability of artificial neural network to extract the relevant underlying patterns in datasets while filtering out the noise of the data when the dataset is large enough (55,000 glass compositions therein). Overall, as illustrated in Fig. 5, machine learning and artificial neural network offer a promising route to predict glass properties as a function of composition while relying only on the analysis of existing datasets, that is, with no physical knowledge prerequisite (i.e., "blind machine learning" [34]).

3.2. Physics-informed composition-property regression models

Although "blind machine learning" and artificial neural network can offer reliable predictions, this approach requires the existent of a large amount of data—which is not always available. In addition, the complex nature of artificial neural network models renders their interpretation challenging, which limits their potential to offer new physical insights. As an alternative route, the concept of "physics-informed machine learning" was recently introduced by Liu et al. [34]. This approach relies on (i) using a simple, analytical model formulation (e.g., a polynomial function) that offers a good interpretability, (ii) linearizing the relationship between inputs and output based on our physical and chemical understanding of the predicted property to increase the propensity of the model for reliable extrapolations, and (iii) identifying relevant reduced-dimensionality descriptors that capture the atomic structure of the glass [34,78]. This approach was recently used to predict the stage I dissolution rate of Na_2O–Al_2O_3–SiO_2 silicate glasses as a function of their composition and pH based on a small dataset (~200 data points) [34].

Fig. 6 presents a comparison between the outcomes of blind and physics-informed machine learning using polynomial regression [34]. In the case of blind machine learning, we find that the optimal model consists of a degree 5 polynomial function. However, as shown in Fig. 6a, this model yields poor predictions as the relative-root-mean-square-error (RRMSE) of the training and test sets are very high, namely, 98% and 731%, respectively [79]. This shows that, in this case, blind machine learning (i.e., the direct prediction of the dissolution rate as a function of composition and pH) requires the use of complex machine learning algorithms (e.g., artificial neural network) and cannot be achieved by simpler, more interpretable models like polynomial regression [36,80].

In contrast, as shown in Fig. 6b, the physics-informed model offers a significantly improved accuracy—with a RRMSE values of 2.32% and 3.77% for the training and test sets, respectively [34]. This was primarily accomplished by using some physical and chemical understanding of the dissolution process of silicate glasses to linearize the relationship between the inputs (i.e., glass composition and pH) and output (i.e., dissolution rate). This greatly decreases the complexity of the model (i.e., polynomial degree 1 as compared to 5 in the case of blind machine learning). In addition, the number of topological constraints per atom (n_c) was introduced as a reduced-dimensionality descriptor that captures how the structure of the glass network controls its dissolution rate [81–86]. This greatly increases the ability of the model to offer some reliable extrapolations far from the initial training set

[34].

Overall, this work suggests that embedding some physical knowledge within machine learning offers a promising route to overcome the tradeoff between accuracy, simplicity, and interpretability (i.e., the degree to which a human can understand the outcome produced by the model [20,24,38])—which are otherwise often mutually exclusive in traditional, blind machine learning models [20,36,54]. Indeed, simple and interpretable models (e.g., polynomial regression) usually offer limited accuracy (see Fig. 6a), whereas more advanced models (e.g., random forest or artificial neural network) offer increased levels of accuracy but often come with higher complexity and lower interpretability (see Fig. 5) [20,36,54]. In general, models that are simpler and more interpretable are highly desirable as (i) simpler models are less likely to overfit small datasets, (ii) simpler models are usually more computationally-efficient, and (iii) more interpretable models are more likely to offer some new insights into the underlying physics governing the relationship between inputs and outputs.

3.3. Composition-property regression models informed by high-throughput simulations

In general, irrespective of the algorithm that is used, the quality of machine learning models depends on the availability of a large body of accurate and consistent data to spans a large compositional domain [31,34]. Since extensive experimental datasets are not always available, high-throughput molecular dynamics (MD) simulations offer a convenient and reliable route to build large, consistent, and accurate datasets of glass properties, which, in turn, can serve as a training set for machine learning algorithms [11,77].

This approach was recently used by Yang et al. to predict the Young's modulus of silicate glasses as a function of their composition [77]. Fig. 7a shows the Young's modulus values E computed by high-throughput MD simulations as a function of composition in the $CaO–Al_2O_3–SiO_2$ glass ternary system [77]. The use of high-throughput MD simulations makes it possible to systematically and homogeneously explore entire compositional domain in an efficient fashion. Importantly, MD simulations offer excellent accuracy and low noise-to-signal ratios, which is key for the use of data-driven modeling. Fig. 7b then shows the prediction of an artificial neural network model trained based on the data present in Fig. 7a [77]. The artificial neural network is found to successfully capture the complex, non-linear evolution of the Young's modulus as a function of composition while filtering out the intrinsic noise of the simulation data. Overall, the model offers an excellent agreement with molecular dynamics data (see Fig. 7c)—with a correlation coefficient R^2 of 0.981 and 0.974 for the training and test sets, respectively. Importantly, the predicted values also show a very good agreement with available experimental data (see Fig. 7d). Note that, although the cooling rate used in MD simulation is significantly higher than experimental ones, computed stiffness values remain fairly unaffected by the cooling rate—as they are mostly governed by the curvature of the interatomic potential [87,88].

These results illustrate the benefits of combining machine learning with high-through MD simulations (i.e., rather than directly relying on available experimental data). Indeed, even for a simple and technologically important system like $CaO–Al_2O_3–SiO_2$ glasses, the number of available experimental stiffness data is fairly limited. Further, most of the data available for this system are clustered in some small regions of the whole compositional domain (namely, pure silica, per-alkaline aluminosilicates, and calcium aluminate glasses). Such clustering of the data is a serious issue as, in turn, available experimental data come with a notable uncertainty—for instance, the Young's modulus of select glasses (at fixed composition) can vary by as much as 20 GPa among different references [32, 89]. As such, the combination of a high level of noise and clustering of the data would not allow machine learning approaches to isolate the “true” trend of the data from their noise. Finally, generating data using MD simulations is faster and cheaper than conducting systematic experiments. Nevertheless, it should be pointed out that, due to some intrinsic limitation of timescale, MD simulations cannot describe the long-term behavior of glasses (e.g., long-term aging or dissolution kinetics). In that regard, various modeling techniques (ranging from physics-based to purely empirical) often needs to be combined to bridge the gap between different timescales [6]. Overall, the combination of physics-based simulations with data-driven machine learning offers a promising route to accelerate the discovery of novel glasses.

3.4. Identification of relevant structural fingerprints

Due to the complex, disordered nature of the atomic structure of glasses, the atoms of the network can exhibit a variety of local environments, which mainly depend on the glass composition and the cooling rate—in contrast with the case of crystals [4]. Such structural complexity makes it challenging to understand how the atomic structure of glasses controls their properties [6,7,102]. Although some properties (e.g., stiffness [88,103] and hardness [10,104]) are largely governed by “intuitive” structural features (e.g., the average coordination number [9,88]), more complex properties (e.g., those that strongly depend on the medium-range order) do not exhibit any correlation with conventional structural metrics [105]. New advanced structural descriptors are required to describe such complex properties (e.g., which cannot be simply described in terms of the average connectivity of the atomic network).

Thanks to its ability to decipher some patterns in complex, multi-dimensional data, machine learning offers a promising route to identify some non-intuitive structural fingerprints that govern glass properties [106]. Recently, Cubuk et al. introduced a classification-based machine learning method to identify some “high-level” structural fingerprints (called “softness”) that control the dynamics of atom rearrangements [102,105–109]. In details, the atomic softness is a highly non-intuitive structural property that is calculated based on the local environment of each atom [106]. This property was determined by classifying each atom as being “soft” (i.e., mobile) or “hard” (i.e., immobile). A large number of systematic structural order descriptors were then computed and used as inputs. A classification model (SVM) was then used to identify the optimal hyperplane within the inputs space that best separates soft from hard atoms (see Fig. 3b). The atomic softness was then defined—for a given atom—as the orthogonal distance between a given position in the inputs space and the hyperplane [107]. As shown in Fig. 8, the probability of atomic rearrangement (P_R) is found to be a logarithmic function of their softness (S) at different temperatures, including into the supercooled liquid regime [105]. Although this approach has thus far been applied to only “toy” model glasses (i.e., Lennard Jones glasses) that may not capture the complex chemistry of more realistic oxide glasses, this work offers some pioneering insights into the linkages between atomic structure and glass properties (dynamics, plasticity, etc.) and paves the way toward the discovery by machine learning of new structural fingerprints that are governing glass properties.

3.5. Machine learning forcefields for glass modeling

As discussed in Section 3.3, MD simulations are an important tool to access the atomic structure of glasses and, thereby, decipher the nature of the relationship between glass composition and properties. However, the reliability of MD (or Monte Carlo) simulations is intrinsically limited by that of the interatomic forcefield that is used, which acts as a bottleneck in glass modeling [15]. To this end, machine learning offers a promising route to develop new accurate interatomic forcefields for glass modeling in an efficient and non-biased fashion [110]. Although various studies have focused on the use of machine learning to develop complex, non-analytical interatomic forcefields, such forcefields present low interpretability and have been largely restricted to simple

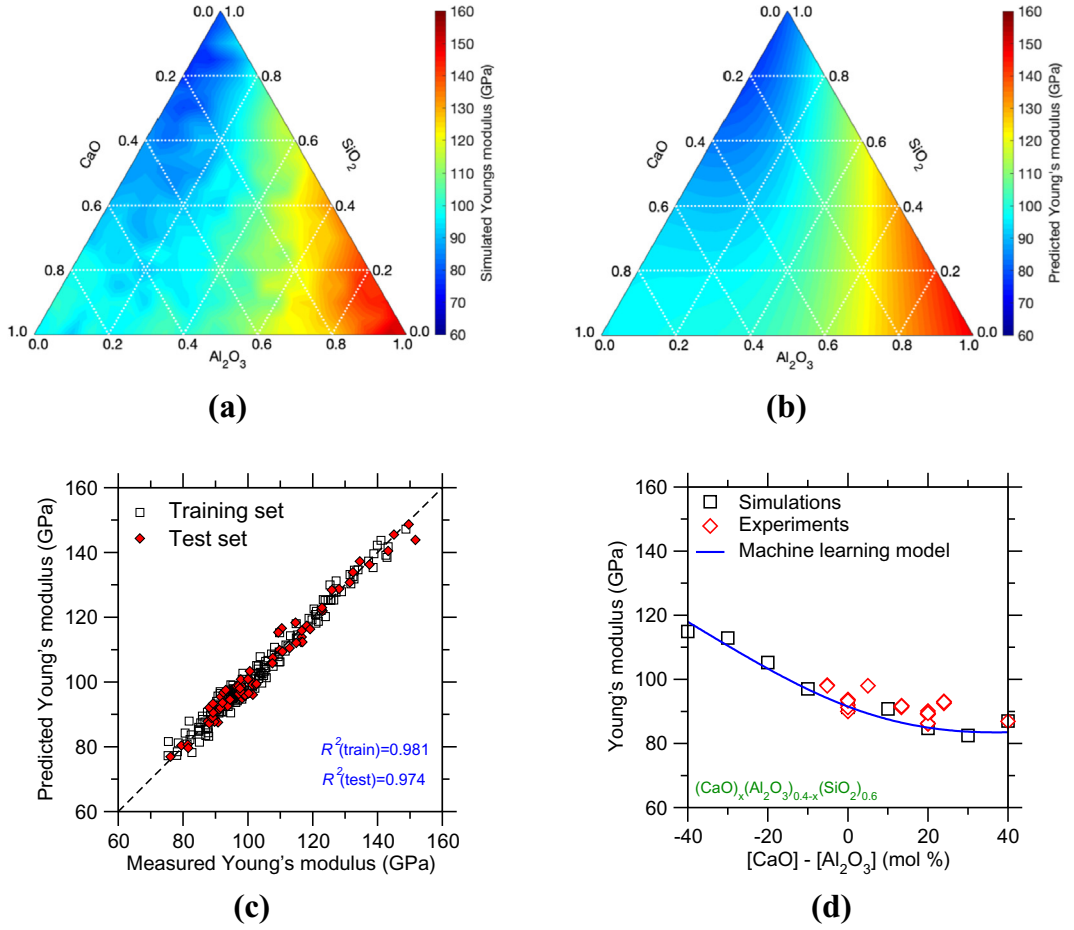

Fig. 7. Ternary diagram showing the Young's modulus values E as a function of composition in the $CaO–Al_2O_3–SiO_2$ glass system (a) computed by high-throughput molecular dynamics (MD) simulations and (b) predicted by artificial neural network (ANN) [77]. (c) Comparison between the Young's modulus values predicted by the ANN model and computed by MD simulations. (d) Comparison between the Young's modulus values computed by MD simulations and predicted by ANN with select available experimental data [89–100] for the series of compositions $(CaO)_x(Al_2O_3)_{40-x}(SiO_2)_{60}$. Note that no experimental data is available for glasses wherein $[CaO] < [Al_2O_3]$ due to the poor glass-forming ability of such compositions [101].

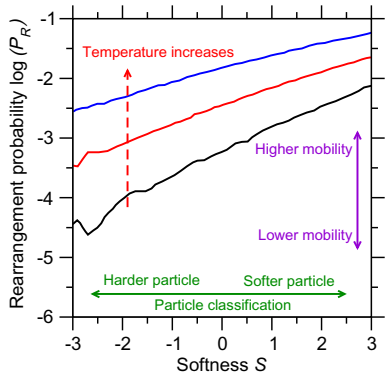

Fig. 8. Probability of atomic rearrangement P_R for select temperatures as the function of the atomic softness—a non-intuitive structural fingerprint identified by classification-based machine learning [105].

monoatomic or diatomic systems thus far [111–114].

On the other hand, empirical forcefields based on analytical forms can offer a realistic description of the atomic structure of silicate glasses [110,115–120]. However, the parameterization of empirical forcefield remains a complex task that often relies on some level of intuition. The parameterization of a forcefield is usually based on the formulation of a cost function that depends on the forcefield parameters [117,119,120].

Recently, Carré et al. introduced a new type of cost function that captures the structural difference between a liquid simulated by ab initio molecular dynamics (i.e., the reference configuration) and that predicted by the forcefield that is to be optimized [120]. The parameterization of the forcefield then consists in identifying the optimal forcefield parameters that minimize the cost function. Traditionally, this step has been conducted by classical minimization algorithms, e.g., steepest gradient descent methods [121]—wherein, starting from a random initial position in the parameter space, one follows the direction of steepest gradient descent in the parameter space until the gradient becomes zero, that is, until a minimum has been found. However, such techniques usually yield some local rather than global minima of functions and, as such, the outcome of the minimization strongly depends on the choice of the initial parameters—which renders the parameterization of forcefield largely biased [119,121].

To overcome these limitations, Liu et al. recently introduced a new forcefield parametrization scheme that combines Gaussian Process Regression and Bayesian optimization [110,122]. The main idea of this method is presented in Fig. 9. Taking glassy silica as an archetypal example, Fig. 9a shows the evolution of the cost function R_χ that is to be minimized as a function of a forcefield parameter (here, the partial charge of Si atoms q_{Si}). The other forcefield parameters are here kept fixed. The cost function R_χ is first interpolated by the GPR method (see Section 2.1 [44]) based on a series of known points, that is, a series of forcefield parameters for which the value of the cost function has been computed. The Figure also shows the uncertainty (95% confidence

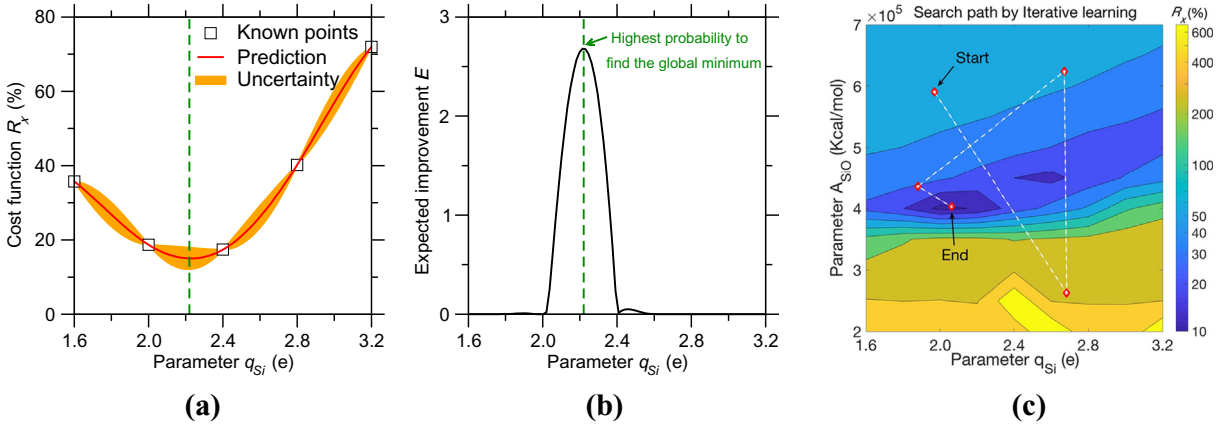

Fig. 9. Illustration of empirical forcefields parametrization using Bayesian optimization and Gaussian Process Regression (GPR) [110]. (a) Cost function R_χ as a function of a forcefield parameter (here, the partial charge of Si atoms q_{Si}). The other forcefield parameters are kept fixed. The cost function R_χ is interpolated by GPR (red line) based on an initial training set comprising 5 data points (i.e., known points, black symbols). The orange area indicates the uncertainty (95% confidence interval) of the prediction. (b) Expected Improvement (EI) function yielded by the Bayesian optimization method, which predicts the set of parameters (here, q_{Si}) that offers the highest probability to find the global minimum of R_χ. (c) Illustration of the iterative parameterization process based on Bayesian optimization. The contour plot shows the cost function R_χ as a function of two select forcefield parameters (q_{Si} and A_{SiO}). The other forcefield parameters are kept fixed. The set of parameters (red diamond) predicted by Bayesian optimization at each iteration is incorporated into the training set, which is used for the next prediction. The white dashed line indicates the path explored by the Bayesian optimization method until the global minimum in the cost function R_χ is identified. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

interval) of the prediction. Based on the GPR predictions and the uncertainty thereof, Bayesian optimization (BO) is then used to determine an optimal set of forcefield parameters that presents the highest probability to yield a minimum value for the cost function R_χ. This is accomplished by using a so-called expected improvement (EI) function (see Fig. 9b) [122], which offers the best tradeoff between "exploitation and exploration," that is, the optimal balance between (i) exploiting the minimum position predicted by GPR and (ii) exploring the parameter space to minimize the uncertainty of the GPR model. The "true" cost function R_χ associated with this optimal set of parameters is then computed by MD and is subsequently incorporated into the training set—which, in turn, refines the GPR model. New optimal forcefield parameters are then iteratively predicted until a satisfactory minimum in the cost function is obtained, that is, when R_χ does not decrease any further. This iterative optimization method is illustrated in Fig. 9c, which shows the path explored by the Bayesian optimization method until the global minimum in the cost function R_χ is identified. This parameterization method is found to yield a forcefield for glassy silicate that offers an unprecedented level of agreement with ab initio simulations [110]. Overall, this work establishes machine learning as a promising route to accelerate the development of new forcefield to model complex, multi-component glasses.

4. Conclusions and future directions

Overall, machine learning techniques offer a unique, largely untapped opportunity to leapfrog current glass design approaches—a process that has thus far remained largely empirical and based on previous experience. When combined with physics-based modeling, machine learning can efficiently and robustly interpolate and extrapolate predictions of glass properties as a function of composition and, hence, drastically accelerate the discovery of new glass formulations with tailored properties and functionalities.

It is worth pointing out that, when adopting machine learning, different properties may come with different challenges and different degrees of complexity. Various criteria can be used to describe the complexity of a given property, e.g.: (i) Does it present a linear or nonlinear dependence on composition? (ii) Is it mostly governed by the short-range order structure of the glass or also sensitive to the medium-range order? (iii) Is it significantly affected by some variations in the thermal history of the glass (e.g., varying cooling rate)? (iv) What is our physical or chemical understanding of the nature of this property? (v) How many existing experimental or simulation data points are available for this property? Clearly, different machine learning algorithms to predict properties with different degrees of complexity—for instance, polynomial regression might be sufficient to predict "simple properties" but more advanced algorithms (e.g., artificial neural network) might be required to model more "complex properties." In addition, predicting more complex properties typically requires larger initial training sets.

Despite these challenges, future applications of machine learning to glass science and engineering are promising and limitless. First, the compositional evolution of virtually all the glass properties can be predicted by machine learning—provided that enough data points are available. To this end, high-throughput atomistic simulations offer a promising route to generate large bodies of consistent, accurate data that can used be as training sets for machine learning approaches. In turn, machine learning optimization techniques offer a unique opportunity to develop new sets of reliable, transferable, and computationally-efficient forcefields for atomistic modeling. In parallel, much progress is still needed to develop new strategies to leverage our existing physical and chemical knowledge of the glassy state to inform machine learning and, hence, overcome some of its intrinsic limitations (e.g., balance between accuracy, complexity, and interpretability). In addition, by excelling at detecting non-intuitive patterns in complex, multi-dimensional datasets, machine learning has the potential to offer some new physical insights into the nature of the glassy state—which have remained hidden thus far due to the complex, disordered, out-of-equilibrium structure of glasses. We postulate that future progress in such approaches will strongly rely on a closer collaboration between different research groups focusing on experiments, theory, simulations, and data analytics. Indeed, successful future applications of machine learning modeling are likely to require closed-loop integrated approaches, wherein (i) experimental or simulation data are used to train machine learning models, (ii) machine learning models are used to pinpoint promising glass compositions, (iii) experiments are conducted to validate these predictions or refined the data-driven models. We hope that the present review will contribute to stimulating the adoption of machine learning techniques in glass science and engineering!

Declaration of Competing Interests

None

Acknowledgments

This work was supported by the National Science Foundation under Grants No. 1762292, 1826420, and 1928538. Part of this research is being performed using funding received from the DOE Office of Nuclear Energy's Nuclear Energy University Program.

References

[1] J.C. Mauro, C.S. Philip, D.J. Vaughn, M.S. Pambianchi, Glass science in the United States: current status and future directions, Int. J. Appl. Glas. Sci. 5 (2014) 2–15, https://doi.org/10.1111/ijag.12058.
[2] J.C. Mauro, E.D. Zanotto, Two centuries of glass research: historical trends, current status, and grand challenges for the future, Int. J. Appl. Glas. Sci. 5 (2014) 313–327, https://doi.org/10.1111/ijag.12087.
[3] E.D. Zanotto, F.A.B. Coutinho, How many non-crystalline solids can be made from all the elements of the periodic table? J. Non-Cryst. Solids 347 (2004) 285–288, https://doi.org/10.1016/j.jnoncrysol.2004.07.081.
[4] A.K. Varshneya, Fundamentals of Inorganic Glasses, Academic Press Inc, 1993.
[5] H. Liu, T. Du, N.M.A. Krishnan, H. Li, M. Bauchy, Topological Optimization of Cementitious Binders: Advances and Challenges, Cement and Concrete Composites, (2018), https://doi.org/10.1016/j.cemconcomp.2018.08.002.
[6] J.C. Mauro, Decoding the glass genome, Curr. Opinion Solid State Mater. Sci. 22 (2018) 58–64, https://doi.org/10.1016/j.cossms.2017.09.001.
[7] 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.
[8] J.C. Mauro, Topological constraint theory of glass, Am. Ceram. Soc. Bull. 90 (2011) 31–37.
[9] 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.
[10] M.M. Smedskjaer, J.C. Mauro, Y. Yue, Prediction of glass hardness using temperature-dependent constraint theory, Phys. Rev. Lett. 105 (2010), https://doi.org/10.1103/PhysRevLett.105.115503.
[11] J.C. Mauro, A. Tandia, K.D. Vargheese, Y.Z. Mauro, M.M. Smedskjaer, Accelerating the design of functional glasses through modeling, Chem. Mater. 28 (2016) 4267–4277, https://doi.org/10.1021/acs.chemmater.6b01054.
[12] 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.
[13] C. Massobrio, J. Du, M. Bernasconi, P.S. Salmon (Eds.), Molecular Dynamics Simulations of Disordered Materials, Springer International Publishing, Cham, 2015, , https://doi.org/10.1007/978-3-319-15675-0.
[14] K. Binder, W. Kob, Glassy Materials and Disordered Solids: An Introduction to their Statistical Mechanics, World Scientific Publishing Company, Hackensack, NJ, 2005.
[15] J. Du, Challenges in molecular dynamics simulations of multicomponent oxide glasses, in: C. Massobrio, J. Du, M. Bernasconi, P.S. Salmon (Eds.), Molecular Dynamics Simulations of Disordered Materials, Springer International Publishing, 2015, pp. 157–180.
[16] L. Huang, J. Kieffer, Challenges in modeling mixed ionic-covalent glass formers, in: C. Massobrio, J. Du, M. Bernasconi, P.S. Salmon (Eds.), Molecular Dynamics Simulations of Disordered Materials, Springer International Publishing, 2015, pp. 87–112, , https://doi.org/10.1007/978-3-319-15675-0_4.
[17] X. Li, W. Song, K. Yang, N.M.A. Krishnan, B. Wang, M.M. Smedskjaer, J.C. Mauro, G. Sant, M. Balonis, M. Bauchy, Cooling rate effects in sodium silicate glasses: bridging the gap between molecular dynamics simulations and experiments, J. Chem. Phys. 147 (2017) 074501, , https://doi.org/10.1063/1.4998611.
[18] M.C. Onbaşlı, A. Tandia, J.C. Mauro, Mechanical and compositional Design of High-Strength Corning Gorilla® Glass, in: W. Andreoni, S. Yip (Eds.), Handbook of Materials Modeling: Applications: Current and Emerging Materials, Springer International Publishing, Cham, 2018, pp. 1–23, , https://doi.org/10.1007/978-3-319-50257-1_100-1.
[19] S.J. Russell, S.J. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, Prentice Hall, 2010.
[20] E. Alpaydin, Introduction to Machine Learning, MIT Press, 2014.
[21] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, ArXiv:1409.1556 [Cs], 2014. http://arxiv.org/abs/1409.1556.
[22] X. Wu, X. Zhu, G. Wu, W. Ding, Data mining with big data, IEEE Trans. Knowl. Data Eng. 26 (2014) 97–107, https://doi.org/10.1109/TKDE.2013.109.
[23] S. Tsugawa, T. Yatabe, T. Hirose, S. Matsumoto, An Automobile with Artificial Intelligence, Proceedings of the 6th International Joint Conference on Artificial Intelligence - Volume 2, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1979, pp. 893–895 http://dl.acm.org/citation.cfm?id=1623050.1623117.
[24] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, 2006.
[25] D.R. Hush, B.G. Horne, Progress in supervised neural networks, IEEE Signal Process. Mag. 10 (1993) 8–39, https://doi.org/10.1109/79.180705.
[26] N.R. Draper, H. Smith, Applied Regression Analysis, John Wiley & Sons, 2014.
[27] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley & Sons, 2012.
[28] H.B. Barlow, Unsupervised learning, Neural Comput. 1 (1989) 295–311, https://doi.org/10.1162/neco.1989.1.3.295.
[29] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Comput. Surv. 31 (1999) 264–323, https://doi.org/10.1145/331499.331504.
[30] G. Cumming, F. Fidler, D.L. Vaux, Error bars in experimental biology, J. Cell Biol. 177 (2007) 7–11, https://doi.org/10.1083/jcb.200611141.
[31] T. Mitchell, B. Buchanan, G. DeJong, T. Dietterich, P. Rosenbloom, A. Waibel, Machine learning, Ann. Rev. Comput. Sci. 4 (1990) 417–433, https://doi.org/10.1146/annurev.cs.04.060190.002221.
[32] A.I. Priven, O.V. Mazurin, Glass property databases: their history, present state, and prospects for further development, Adv. Mater. Res. 39–40 (2008) 145–150, https://doi.org/10.4028/www.scientific.net/AMR.39-40.145.
[33] R. Chrisley, Embodied artificial intelligence, Artif. Intell. 149 (2003) 131–150, https://doi.org/10.1016/S0004-3702(03)00055-9.
[34] H. Liu, T. Zhang, N.M.A. Krishnan, M.M. Smedskjaer, J.V. Ryan, S. Gin, M. Bauchy, Physics-Informed Machine Learning: Predicting the Stage I Dissolution Kinetics of Silicate Glasses, Npj Materials Degradation, 2019.
[35] A.L. Pomerantsev, Confidence intervals for nonlinear regression extrapolation, Chemom. Intell. Lab. Syst. 49 (1999) 41–48, https://doi.org/10.1016/S0169-7439(99)00026-X.
[36] 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.
[37] D.R. Cassar, A.C.P.L.F. de Carvalho, E.D. Zanotto, Predicting glass transition temperatures using neural networks, Acta Materialia 159 (2018) 249–256, https://doi.org/10.1016/j.actamat.2018.08.022.
[38] T. Lookman, F. Alexander, K. Rajan, Information Science for Materials Discovery and Design, Springer, Berlin Heidelberg, New York, 2015.
[39] G.A.F. Seber, A.J. Lee, Linear Regression Analysis, John Wiley & Sons, 2012.
[40] Yu.N. Subbotin, Piecewise-polynomial (spline) interpolation, Math. Notes Acad. Sci. USSR 1 (1967) 41–45, https://doi.org/10.1007/BF01221723.
[41] H.J. Motulsky, L.A. Ransnas, Fitting curves to data using nonlinear regression: a practical and nonmathematical review, FASEB J. 1 (1987) 365–374, https://doi.org/10.1096/fasebj.1.5.3315805.
[42] W. Härdle, Applied Nonparametric Regression, Cambridge University Press, 1990.
[43] N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, Am. Stat. 46 (1992) 175–185, https://doi.org/10.1080/00031305.1992.10475879.
[44] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, 3. print MIT Press, Cambridge, Mass, 2008.
[45] Y.L. Tong, The Multivariate Normal Distribution, Springer-Verlag, New York, 1990.
[46] S. Bishnoi, S. Singh, R. Ravinder, M. Bauchy, N.N. Gosvami, H. Kodamana, N.M.A. Krishnan, Predicting Young's Modulus of Glasses with Sparse Datasets using Machine Learning, ArXiv:1902.09776 [Cond-Mat], 2019. http://arxiv.org/abs/1902.09776.
[47] K. Mohiuddin, J. Mao, A.K. Jain, Artificial neural networks: a tutorial, Computer. 29 (1996) 31–44.
[48] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Disc. 2 (1998) 121–167, https://doi.org/10.1023/A:1009715923555.
[49] A. Liaw, M. Wiener, Classification and regression by randomForest, R News. 2 (2002) 18–22.
[50] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (2001) 1189–1232.
[51] N. Cristianini, J. Shawe-Taylor, D. of C.S.R.H.J. Shawe-Taylor, J. (Royal H.S.-T. London) University of, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge University Press, 2000.
[52] R.M. Balabin, E.I. Lomakina, Support vector machine regression (SVR/LS-SVM)—an alternative to neural networks (ANN) for analytical chemistry? Comparison of nonlinear methods on near infrared (NIR) spectroscopy data, Analyst. 136 (2011) 1703–1712, https://doi.org/10.1039/C0AN00387E.
[53] C. Wang, S.S. Venkatesh, J.S. Judd, Optimal stopping and effective machine complexity in learning, in: J.D. Cowan, G. Tesauro, J. Alspector (Eds.), Advances in Neural Information Processing Systems 6, Morgan-Kaufmann, 1994, pp. 303–310.
[54] E. Aragones, I. Gilboa, A. Postlewaite, D. Schmeidler, Accuracy vs. simplicity: a complex trade-off, SSRN Electron. J. (2002), https://doi.org/10.2139/ssrn.332382.
[55] J. Lever, M. Krzywinski, N. Altman, Model selection and overfitting: points of significance, Nat. Methods 13 (2016) 703–704, https://doi.org/10.1038/nmeth.3968.
[56] Y. Bengio, Y. Grandvalet, No unbiased estimator of the variance of K-fold cross-validation, J. Mach. Learn. Res. 5 (2004) 1089–1105.
[57] H. Zou, T. Hastie, Regularization and variable selection via the elastic net, J. R. Stat. Soc. 67 (2005) 301–320, https://doi.org/10.1111/j.1467-9868.2005.00503.x.
[58] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. R. Stat. Soc. Ser. B Methodol. 58 (1996) 267–288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.
[59] A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, Technometrics. 12 (1970) 55–67, https://doi.org/10.1080/00401706.

1970.10488634.
[60] M.W. Gardner, S.R. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, Atmos. Environ. 32 (1998) 2627–2636, https://doi.org/10.1016/S1352-2310(97)00447-0.
[61] T. Hastie, S. Rosset, J. Zhu, H. Zou, Multi-class AdaBoost, Stat. Interface. 2 (2009) 349–360, https://doi.org/10.4310/SII.2009.v2.n3.a8.
[62] D. Wunsch, Survey of clustering algorithms, IEEE Trans. Neural Netw. 16 (2005) 645–678, https://doi.org/10.1109/TNN.2005.845141.
[63] A.K. Jain, Data clustering: 50 years beyond K-means, Pattern Recogn. Lett. 31 (2010) 651–666, https://doi.org/10.1016/j.patrec.2009.09.011.
[64] T.S. Madhulatha, An Overview on Clustering Methods, ArXiv:1205.1117 [Cs], 2012. http://arxiv.org/abs/1205.1117.
[65] A. Likas, N. Vlassis, J.J. Verbeek, The global k-means clustering algorithm, Pattern Recogn. 36 (2003) 451–461, https://doi.org/10.1016/S0031-3203(02)00060-2.
[66] T.M. Kodinariya, P.R. Makwana, Review on determining number of cluster in K-means clustering, Int. J. Adv. Res. Comput. Sci. Manage. Stud. 1 (2013) 6.
[67] P. Bholowalia, EBK-means: a clustering technique based on elbow method and K-means in WSN, Int. J. Comput. Appl. 105 (2014) 17–24.
[68] P.S. Bradley, U.M. Fayyad, Refining Initial Points for K-Means Clustering, Morgan Kaufmann, 1998, pp. 91–99.
[69] S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, 2014 Science and Information Conference, 2014, pp. 372–378, , https://doi.org/10.1109/SAI.2014.6918213.
[70] M.I. Jordan, T.M. Mitchell, Machine learning: trends, perspectives, and prospects, Science. 349 (2015) 255–260, https://doi.org/10.1126/science.aaa8415.
[71] I. Jolliffe, Principal component analysis, Encyclopedia of Statistics in Behavioral Science, American Cancer Society, 2005, , https://doi.org/10.1002/0470013192.bsa501.
[72] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature. 401 (1999) 788–791, https://doi.org/10.1038/44565.
[73] M. Li, B. Yuan, 2D-LDA: a statistical linear discriminant analysis for image matrix, Pattern Recogn. Lett. 26 (2005) 527–532, https://doi.org/10.1016/j.patrec.2004.09.007.
[74] L. van der Maaten, E. Postma, J. van den Herik, Dimensionality Reduction: A Comparative Review, http://www.math.chalmers.se/Stat/Grundutb/GU/MSA220/S18/DimRed2.pdf.
[75] D.S. Brauer, C. Rüssel, J. Kraft, Solubility of glasses in the system P2O5–CaO–MgO–Na2O–TiO2: experimental and modeling using artificial neural networks, J. Non-Cryst. Solids 353 (2007) 263–270, https://doi.org/10.1016/j.jnoncrysol.2006.12.005.
[76] J. Lee Rodgers, W.A. Nicewander, Thirteen ways to look at the correlation coefficient, Am. Stat. 42 (1988) 59–66, https://doi.org/10.1080/00031305.1988.10475524.
[77] K. Yang, X. Xu, B. Yang, B. Cook, H. Ramos, M. Bauchy, Prediction of Silicate Glasses' Stiffness by High-Throughput Molecular Dynamics Simulations and Machine Learning, ArXiv:1901.09323 [Cond-Mat, Physics:Physics], 2019. http://arxiv.org/abs/1901.09323.
[78] J.D. Vienna, J.J. Neeway, J.V. Ryan, S.N. Kerisit, Impacts of glass composition, pH, and temperature on glass forward dissolution rate, Npj Mater. Degrad. 2 (2018) 22, https://doi.org/10.1038/s41529-018-0042-5.
[79] M.-F. Li, X.-P. Tang, W. Wu, H.-B. Liu, General models for estimating daily global solar radiation for different solar radiation zones in mainland China, Energy Convers. Manag. 70 (2013) 139–148, https://doi.org/10.1016/j.enconman.2013.03.004.
[80] P. Frugier, C. Martin, I. Ribet, T. Advocat, S. Gin, The effect of composition on the leaching of three nuclear waste glasses: R7T7, AVM and VRZ, J. Nucl. Mater. 346 (2005) 194–207, https://doi.org/10.1016/j.jnucmat.2005.06.023.
[81] T. Oey, A. Kumar, I. Pignatelli, Y. Yu, N. Neithalath, J.W. Bullard, M. Bauchy, G. Sant, Topological controls on the dissolution kinetics of glassy aluminosilicates, J. Am. Ceram. Soc. 100 (2017) 5521–5527, https://doi.org/10.1111/jace.15122.
[82] I. Pignatelli, A. Kumar, M. Bauchy, G. Sant, Topological control on silicates' dissolution kinetics, Langmuir. 32 (2016) 4434–4439, https://doi.org/10.1021/acs.langmuir.6b00359.
[83] T. Oey, Y.-H. Hsiao, E. Callagon, B. Wang, I. Pignatelli, M. Bauchy, G.N. Sant, Rate controls on silicate dissolution in cementitious environments, RILEM Tech. Lett. 2 (2017) 67–73, https://doi.org/10.21809/rilemtechlett.2017.35.
[84] T. Oey, K.F. Frederiksen, N. Mascaraque, R. Youngman, M. Balonis, M.M. Smedskjaer, M. Bauchy, G. Sant, The role of the network-modifier's field-strength in the chemical durability of aluminoborate glasses, J. Non-Cryst. Solids 505 (2019) 279–285, https://doi.org/10.1016/j.jnoncrysol.2018.11.019.
[85] N. Mascaraque, M. Bauchy, M.M. Smedskjaer, Correlating the network topology of oxide glasses with their chemical durability, J. Phys. Chem. B 121 (2017) 1139–1147, https://doi.org/10.1021/acs.jpcb.6b11371.
[86] N. Mascaraque, M. Bauchy, J.L.G. Fierro, S.J. Rzoska, M. Bockowski, M.M. Smedskjaer, Dissolution kinetics of hot compressed oxide glasses, J. Phys. Chem. B 121 (2017) 9063–9072, https://doi.org/10.1021/acs.jpcb.7b04535.
[87] J.F. Lutsko, Generalized expressions for the calculation of elastic constants by computer simulation, J. Appl. Phys. 65 (1989) 2991–2997, https://doi.org/10.1063/1.342716.
[88] Rouxel Tanguy, Elastic properties and short-to medium-range order in glasses, J. Am. Ceram. Soc. 90 (2007) 3019–3039, https://doi.org/10.1111/j.1551-2916.2007.01945.x.
[89] L.-G. Hwa, K.-J. Hsieh, L.-C. Liu, Elastic moduli of low-silica calcium aluminosilicate glasses, Mater. Chem. Phys. 78 (2003) 105–110, https://doi.org/10.1016/S0254-0584(02)00331-0.
[90] R.J. Eagan, J.C. Swearekgen, Effect of composition on the mechanical properties of Aluminosilicate and borosilicate glasses, J. Am. Ceram. Soc. 61 (1978) 27–30, https://doi.org/10.1111/j.1151-2916.1978.tb09222.x.
[91] C. Ecolivet, P. Verdier, Proprietes elastiques et indices de refraction de verres azotes, Mater. Res. Bull. 19 (1984) 227–231, https://doi.org/10.1016/0025-5408(84)90094-1.
[92] S. Inaba, S. Todaka, Y. Ohta, K. Morinaga, Equation for estimating the young’s modulus, shear modulus and Vickers hardness of aluminosilicate glasses, J. Jpn. Inst. Metals 64 (2000) 177–183, https://doi.org/10.2320/jinstmet1952.64.3_177.
[93] S. Inaba, S. Oda, K. Morinaga, Equation for estimating the thermal diffusivity, specific heat and thermal conductivity of oxide glasses, J. Jpn. Inst. Metals 65 (2001) 680–687, https://doi.org/10.2320/jinstmet1952.65.8_680.
[94] C. Weigel, C. Le Losq, R. Vialla, C. Dupas, S. Clément, D.R. Neuville, B. Rufflé, Elastic moduli of XAlSiO4 aluminosilicate glasses: effects of charge-balancing cations, J. Non-Cryst. Solids 447 (2016) 267–272, https://doi.org/10.1016/j.jnoncrysol.2016.06.023.
[95] J. Rocherulle, C. Ecolivet, M. Poulain, P. Verdier, Y. Laurent, Elastic moduli of oxynitride glasses: extension of Makishima and Mackenzie's theory, J. Non-Cryst. Solids 108 (1989) 187–193, https://doi.org/10.1016/0022-3093(89)90582-6.
[96] M. Yamane, M. Okuyama, Coordination number of aluminum ions in alkali-free alumino-silicate glasses, J. Non-Cryst. Solids 52 (1982) 217–226, https://doi.org/10.1016/0022-3093(82)90297-6.
[97] S. Sugimura, S. Inaba, H. Abe, K. Morinaga, Compositional dependence of mechanical properties in aluminosilicate, borate and phosphate glasses, J. Ceram. Soc. Jpn. 110 (2002) 1103–1106, https://doi.org/10.2109/jcersj.110.1103.
[98] T.M. Gross, M. Tomozawa, A. Koike, A glass with high crack initiation load: role of fictive temperature-independent mechanical properties, J. Non-Cryst. Solids 355 (2009) 563–568, https://doi.org/10.1016/j.jnoncrysol.2009.01.022.
[99] I. Yasui, F. Utsuno, Material design of glasses based on database – INTERGLAD, in: M. Doyama, J. Kihara, M. Tanaka, R. Yamamoto (Eds.), Computer Aided Innovation of New Materials II, Elsevier, Oxford, 1993, pp. 1539–1544, , https://doi.org/10.1016/B978-0-444-89778-7.50147-X.
[100] N.P. Bansal, R.H. Doremus, Handbook of Glass Properties, Elsevier, 2013.
[101] J.E. Shelby, Formation and properties of calcium aluminosilicate glasses, J. Am. Ceram. Soc. 68 (1985) 155–158, https://doi.org/10.1111/j.1151-2916.1985.tb09656.x.
[102] E.D. Cubuk, R.J.S. Ivancic, S.S. Schoenholz, D.J. Strickland, A. Basu, Z.S. Davidson, J. Fontaine, J.L. Hor, Y.-R. Huang, Y. Jiang, N.C. Keim, K.D. Koshigan, J.A. Lefever, T. Liu, X.-G. Ma, D.J. Magagnosc, E. Morrow, C.P. Ortiz, J.M. Rieser, A. Shavit, T. Still, Y. Xu, Y. Zhang, K.N. Nordstrom, P.E. Arratia, R.W. Carpick, D.J. Durian, Z. Fakhraai, D.J. Jerolmack, D. Lee, J. Li, R. Riggleman, K.T. Turner, A.G. Yodh, D.S. Gianola, A.J. Liu, Structure-property relationships from universal signatures of plasticity in disordered solids, Science 358 (2017) 1033–1037, https://doi.org/10.1126/science.aai8830.
[103] K. Philipps, R.P. Stoffel, R. Dronskowski, R. Conradt, Experimental and theoretical investigation of the elastic moduli of silicate glasses and crystals, Front. Mater. 4 (2017), https://doi.org/10.3389/fmats.2017.00002.
[104] M. Bauchy, M.J.A. Qomi, C. Bichara, F.-J. Ulm, R.J.-M. Pellenq, Rigidity transition in materials: hardness is driven by weak atomic constraints, Phys. Rev. Lett. 114 (2015) 125502, https://doi.org/10.1103/PhysRevLett.114.125502.
[105] S.S. Schoenholz, E.D. Cubuk, D.M. Sussman, E. Kaxiras, A.J. Liu, A structural approach to relaxation in glassy liquids, Nat. Phys. 12 (2016) 469–471, https://doi.org/10.1038/nphys3644.
[106] E.D. Cubuk, S.S. Schoenholz, E. Kaxiras, A.J. Liu, Structural properties of defects in glassy liquids, J. Phys. Chem. B 120 (2016) 6139–6146, https://doi.org/10.1021/acs.jpcb.6b02144.
[107] E.D. Cubuk, S.S. Schoenholz, J.M. Rieser, B.D. Malone, J. Rottler, D.J. Durian, E. Kaxiras, A.J. Liu, Identifying structural flow defects in disordered solids using machine-learning methods, Phys. Rev. Lett. 114 (2015), https://doi.org/10.1103/PhysRevLett.114.108001.
[108] D.M. Sussman, S.S. Schoenholz, E.D. Cubuk, A.J. Liu, Disconnecting structure and dynamics in glassy thin films, PNAS. 114 (2017) 10601–10605, https://doi.org/10.1073/pnas.1703927114.
[109] X. Ma, Z.S. Davidson, T. Still, R.J.S. Ivancic, S.S. Schoenholz, A.J. Liu, A.G. Yodh, Heterogeneous activation, local structure, and softness in supercooled colloidal liquids, Phys. Rev. Lett. 122 (2019), https://doi.org/10.1103/PhysRevLett.122.028001.
[110] H. Liu, Z. Fu, Y. Li, N.F.A. Sabri, M. Bauchy, Machine Learning Forcefield for Silicate Glasses, ArXiv:1902.03486 [Cond-Mat], 2019. http://arxiv.org/abs/1902.03486.
[111] A.P. Bartók, J. Kermode, N. Bernstein, G. Csányi, Machine learning a general-purpose interatomic potential for silicon, Phys. Rev. (2018), https://doi.org/10.1103/PhysRevX.8.041048 X. 8.
[112] V.L. Deringer, G. Csányi, Machine learning based interatomic potential for amorphous carbon, Phys. Rev. B 95 (2017) 094203, , https://doi.org/10.1103/PhysRevB.95.094203.
[113] P. Rowe, G. Csányi, D. Alfè, A. Michaelides, Development of a machine learning potential for graphene, Phys. Rev. B. 97 (2018), https://doi.org/10.1103/PhysRevB.97.054303.
[114] M. Hellström, J. Behler, Neural network potentials in materials modeling, in: W. Andreoni, S. Yip (Eds.), Handbook of Materials Modeling, Springer International Publishing, Cham, 2018, pp. 1–20, , https://doi.org/10.1007/978-3-319-42913-7_56-1.
[115] 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.

[116] L. Deng, J. Du Development of boron oxide potentials for computer simulations of multicomponent oxide glasses, J. Am. Ceram. Soc.. doi:https://doi.org/10.1111/jace.16082.
[117] S. Sundararaman, L. Huang, S. Ispas, W. Kob, New optimization scheme to obtain interaction potentials for oxide glasses, J. Chem. Phys. 148 (2018) 194504, https://doi.org/10.1063/1.5023707.
[118] M. Wang, N.M. Anoop Krishnan, B. Wang, M.M. Smedskjaer, J.C. Mauro, M. Bauchy, A new transferable interatomic potential for molecular dynamics simulations of borosilicate glasses, J. Non-Cryst. Solids 498 (2018) 294–304, https://doi.org/10.1016/j.jnoncrysol.2018.04.063.
[119] A. Carré, S. Ispas, J. Horbach, W. Kob, Developing empirical potentials from ab initio simulations: the case of amorphous silica, Comput. Mater. Sci. 124 (2016) 323–334, https://doi.org/10.1016/j.commatsci.2016.07.041.
[120] A. Carré, J. Horbach, S. Ispas, W. Kob, New fitting scheme to obtain effective potential from Car-Parrinello molecular-dynamics simulations: application to silica, EPL. 82 (2008) 17001, https://doi.org/10.1209/0295-5075/82/17001.
[121] J.R. Shewchuk, An Introduction to the Conjugate Gradient Method without the Agonizing Pain, https://www.cs.cmu.edu/~quake-papers/painless-conjugate-gradient.pdf, (1994).
[122] P.I. Frazier, J. Wang, Bayesian optimization for materials design, Information Science for Materials Discovery and Design, Springer, Cham, 2016, pp. 45–75, , https://doi.org/10.1007/978-3-319-23871-5_3.