



Effective subgrouping enhances machine learning prediction in complex materials science phenomena: Inoue's subgrouping in discovering bulk metallic glasses

Guannan Liu^a, Sungwoo Sohn^a, Corey S. O'Hern^{a,b,c}, Anna C. Gilbert^{d,e}, Jan Schroers^{a,*}

^a Department of Mechanical Engineering and Materials Science, Yale University, New Haven, CT 06520, USA

^b Department of Applied Physics, Yale University, New Haven, Connecticut 06520, USA

^c Department of Physics, Yale University, New Haven, Connecticut 06520, USA

^d Department of Electrical Engineering, Yale University, New Haven, Connecticut 06520, USA

^e Department of Statistics & Data Science, Yale University, New Haven, Connecticut 06520, USA

ARTICLE INFO

Keywords:

Machine learning
Materials discovery
Materials informatics
Metallic glass
Glass-forming ability

ABSTRACT

Addressing complex materials science problems through machine learning (ML) is challenging. A primary reason for the challenge is that the underlying mechanisms may vary within the considered problem space. To quantify this, we divide alloy data into subgroups and construct ML models to predict metallic glass formation. We discover that subgrouping guided by physical insights into the problem leads to significantly higher prediction accuracy. Specifically, when applying Inoue's subgrouping, models specific to subgroups outperform those trained on the entire dataset. Moreover, our analysis uncovers distinct mechanisms and contributing factors that control the glass-forming ability in different subgroups, shedding light on the diverse nature of this phenomenon. Statistical methods for subgrouping prove less effective and constrained when compared to physics-informed subgrouping. Our results underscore the importance of leveraging physical insights for effective subgrouping or precise feature representation, to guide ML strategies when tackling complex materials science problems. Such an integrated approach has the potential to unlock new insights into material composition-property relationships and accelerate materials discovery in a wide range of applications beyond metallic glass formation.

1. Introduction

Complex materials science problems and phenomena often involve a large number of atoms, which are usually many orders of magnitude too large to be addressed by *ab initio* calculations. Examples of such complex materials science problems include the prediction of liquidus temperature of an alloy [1–3], the viscosity of a liquid [4,5], the plastic region of the stress-strain curve [6], microstructure and microstructure evolution and the resulting properties [7–11], and the glass forming ability (GFA) [12–15] of an alloy.

Particularly, bulk metallic glass (BMG) formation has raised significant scientific and technological interest [12,13,16,17]. The technological interest originates from their superb properties and property combinations [18–26] and from their unique-for-metals processability [27]. Scientifically, BMGs give rise to a convenient spatial and temporal window to study the structure of glass and liquids [28,29] and their

structure-property relationships [30–32]. Generally, as BMG formation is a complex process, it has been difficult to predict GFA of alloys.

For such complex problems, empirical rules, model descriptions, and guiding principles have been developed to help understand and predict material behavior and properties and also to guide materials discovery. Examples include the Hume-Rothery rules for solid solution formation [33,34], the Hall-Petch relationship for grain boundary strengthening [35,36], Turnbull's reduced glass transition temperature criteria for bulk glass formation [37], correlation between elastic constants and plasticity in metallic glasses [18,22,38], and Inoue's rules for designing BMGs [13]. Obviously, to apply broadly across the wide range of a complex problem, such rules must be generalized. Thereby, they can describe general trends [12,16,38,39], but at the cost of accounting for material-specific behavior. This has led to refinement and further specification of rules [40–44]. For example, Inoue suggested that more accurate design criteria for BMGs are possible when subgrouping alloys

* Corresponding author.

E-mail address: jan.schroers@yale.edu (J. Schroers).

<https://doi.org/10.1016/j.actamat.2023.119590>

Received 9 October 2023; Received in revised form 29 November 2023; Accepted 7 December 2023

Available online 13 December 2023

1359-6454/© 2023 The Authors. Published by Elsevier Ltd on behalf of Acta Materialia Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

based on their chemical characteristics which are reflected in the location of the alloy constituents in the periodic table [45]. However, thus far, validating such empirical rules or subgrouping has remained challenging as it requires large amounts of consistent and high-quality data. It is obvious that the availability of effective rules guiding metallurgy is of paramount importance when considering the vast composition and processing space available for alloys [46].

In this study, we employ machine learning (ML) strategies to quantify the effectiveness of such rules. Specifically, we apply ML to a comprehensive database of alloys to validate Inoue's empirical rules and subgrouping for metallic glasses [45]. For this, we divide the database into seven subgroups according to Inoue's subgrouping criteria. Subsequently, we build individual random forest ML models for each subgroup of data, using physics-informed features that have been previously confirmed as effective [14]. When predicting into held-out test datasets, we find that the models trained on the same subgroup of data perform better than the models trained on all available data, suggesting that physics-informed subgrouping is more effective at enhancing prediction accuracy than increasing data quantity.

A powerful test of the subgrouping approach is to use models trained on each subgroup to predict into other subgroups, and compare these predictions with predictions where training and testing is performed within the same subgroup. Our results show that predictions within the same subgroup are significantly more accurate than predictions into other subgroups. To calibrate this result, we compare the above results with the predictions generated by models trained on randomly grouped data, assuming no domain-specific knowledge regarding subgrouping. Specifically, we conduct a parallel experiment where we randomly divide all data into seven "subgroups", each equivalent in size to Inoue's seven subgroups and repeat the same process of ML training and prediction. From this benchmark experiment, we observe no significant difference between within-group predictions and out-of-group predictions, while the accuracy of within-group predictions is lower than that using Inoue's subgrouping, revealing the effectiveness of Inoue's subgrouping strategies that are physics informed.

To understand the underlying origin of Inoue's subgrouping, we identify and compare the feature importance across ML models trained on different subgroups. We find that feature importance varies significantly across groups, indicating that motifs for glass formation vary among these Inoue's subgroups.

Overall, our findings suggest that more specific physical insights are needed to develop effective ML models for complex materials science problems. Effective subgrouping solely through data science strategies is generally not feasible due to the astronomical vastness of the parameter space of complex materials science problems and the non-physical approximations that all such data science strategies employ. More accurate ML models can be achieved either by constructing physics-informed features based on human insights [14] or dividing the data into subgroups that follow the same underlying physics, as proposed in this study. Such approaches can be applied to other complex materials science problems beyond metallic glass formation. Additionally, our findings suggest that the effectiveness of empirical rules as guiding principles can be tested, and mechanistic insights can be revealed using ML strategies, potentially leading to the development of more accurate and generalizable rules. Further investigation into the underlying mechanisms of subgrouping could also reveal new insights into the relationship between material composition and properties.

2. Methods

2.1. Data collection and subgrouping

We utilize a database compiled by Liu et al. [14] composed of essentially all experimentally reported data from the Landolt-Bornstein Handbook on "Nonequilibrium Phase Diagrams of Ternary Amorphous Alloys" [47], and peer-reviewed literature on GFA of alloys. We labeled

alloys as either BMG formers (critical cooling rate $R_c < 10^3$ K/s) or non-glass formers ($R_c > 10^6$ K/s), resulting in a database of 2740 unique alloy compositions, with 1027 BMG and 1713 non-glass formers. The database contains 55 distinct elements and considers alloys ranging from binary to octonary alloys, with ternary alloys forming the majority of the alloys.

We used the classification system proposed by Inoue [45] to group alloys into seven distinct subgroups based on their constituent elements (Fig. 1). This classification system groups alloys using the atomic size difference, heat of mixing, and period of the constituent elements in the periodic table. The seven subgroups from Group 1 (G1) to Group 7 (G7) are as follows:

1. ETM/Ln-LTM/BM-Al/Ga
2. ETM/Ln-LTM/BM-Metalloid
3. Al/Ga-LTM/BM-Metalloid
4. IIA-ETM/Ln-LTM/BM
5. LTM/BM-Metalloid
6. ETM/Ln-LTM/BM
7. IIA-LTM/BM

where ETM, Ln, LTM, BM and IIA refer to early transition, lanthanide, late transition, group IIIB-IVB, and group IIA-group metals, respectively.

2.2. Feature construction: physics-informed features

The identification and construction of features that represent the property of interest of an alloy is critical for building effective and interpretable ML models. Here, the property of interest is the GFA of the alloy. Features, summarized in feature vectors, can be defined as a set of quantitative and qualitative attributes that describe the alloy for the property of interest and are the building blocks of ML models. Each alloy composition and label, i.e., BMG or non-glass, corresponds to a feature vector. Physical insights into the problem can be utilized to choose or to construct features more effectively [14]. In our study, we select and build features based on their ability to represent the underlying atomic interactions in the alloy governing GFA. The characteristics of an alloy to be a BMG former are [13]: a) a composition close to a deep eutectic, b) atomic size difference of larger than 12 %, c) a large negative heat of mixing among at least two constituent elements. These empirical rules reflect the established understanding of an alloy's characteristics to form a BMG, which have been also suggested by Inoue [13]. To represent these rules by properties that are *a priori* known, we construct the following features:

1. Liquidus temperature of the alloy, T_{alloy} . To determine T_{alloy} for a general multicomponent alloy, we first break down the alloy into all possible binary combinations. The liquidus temperature for each binary system is readily available, such as in the ASM Alloy Phase Diagram Database. The liquidus temperature of the alloy, T_{alloy} , is constructed using the ratio of these binary combinations. We extrapolate T_{alloy} by calculating it from the liquidus temperatures of constituent binary pairs, denoted as T_{AB} for the composition $A_{\frac{a}{a+b}}B_{\frac{b}{a+b}}$. For example, in the case of a ternary alloy $A_aB_bC_c$, T_{alloy} is calculated as follows:

$$T_{\text{alloy}} = \frac{(a+b) \times T_{AB} + (a+c) \times T_{AC} + (b+c) \times T_{BC}}{2 \times (a+b+c)} \quad (1)$$

Note that for the ternary alloy A-B-C system, a , b , and c are the compositions of elements A, B, and C, respectively.

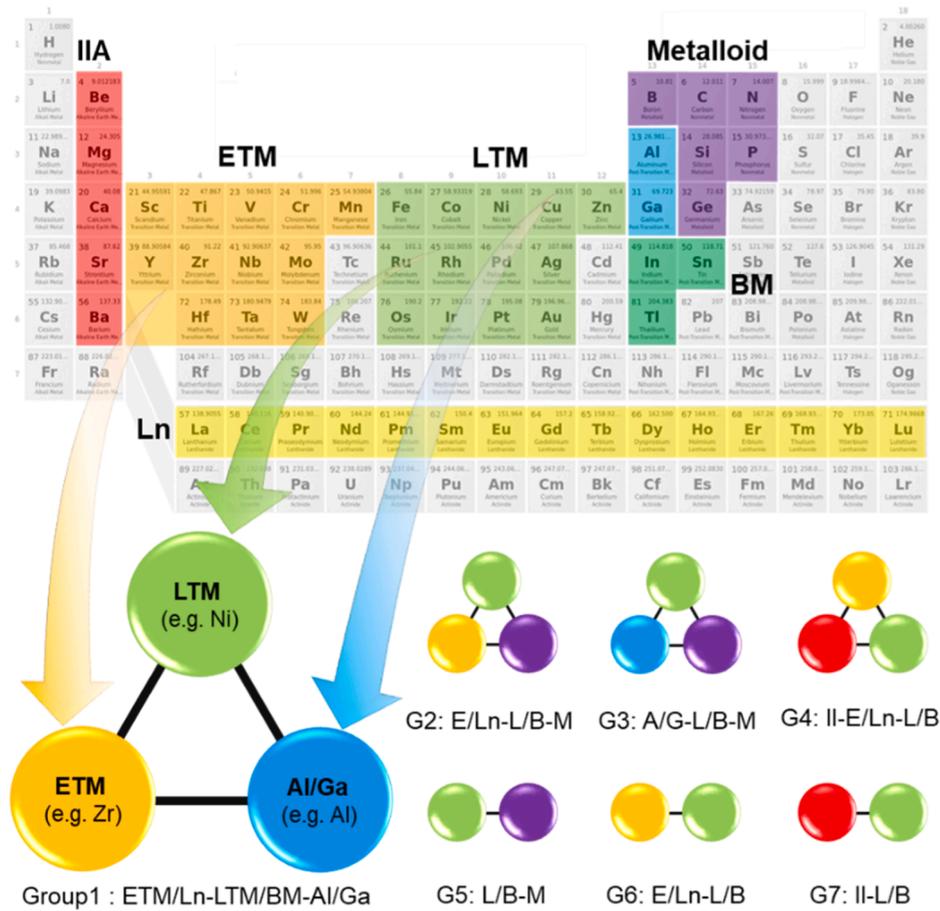


Fig. 1. Inoue's subgrouping of BMGs. The seven subgroups are: (G1) ETM/Ln-LTM/BM-Al/Ga, (G2) ETM/Ln-LTM/BM-Metalloid, (G3) Al/Ga-LTM/BM-Metalloid, (G4) IIA-ETM/Ln-LTM/BM, (G5) LTM/BM-Metalloid, (G6) ETM/Ln-LTM/BM, and (G7) IIA-LTM/BM, where ETM, Ln, LTM, BM and IIA refer to early transition, lanthanide, late transition, group IIIB–IVB, and group IIA-group metals, respectively.

2. Liquidus temperature reduction, T_r : To determine the reduction in liquidus temperature, we normalize T_{alloy} by the mean liquidus temperature among the constituent elements, denoted as T_{mean} . T_r is a dimensionless ratio, ranging from 1 (indicating no reduction) to 0 (indicating total reduction). For the ternary alloy $A_aB_bC_c$, T_{mean} is calculated as $T_A \times a + T_B \times b + T_C \times c$. T_r is then expressed as:

$$T_r = \frac{T_{\text{alloy}}}{T_{\text{mean}}} \quad (2)$$

3. Atomic size difference, δ :

$$\delta = 100\% * \sqrt{\sum_i x_i (1 - r_i/\bar{r})^2}, \bar{r} = \sum_i x_i r_i, \quad (3)$$

where r_i is the atomic radius of the constituent element, and x_i is the atomic fraction of the element. Here, \bar{r} is the mean atomic radius among the constituent elements.

4. Atomic size ratio ϕ :

$$\phi = r_{\text{max}} / r_{\text{min}}, \quad (4)$$

where r_{max} represents the largest atomic radius among the constituent elements, while r_{min} is the smallest atomic radius among the constituent elements.

5. Atomic size range Δr :

$$\Delta r = r_{\text{max}} - r_{\text{min}}, \quad (5)$$

where r_{max} is the largest atomic radius among the constituent elements, and r_{min} is the smallest atomic radius among the constituent elements.

6. Maximum heat of mixing ΔH_{max} : We first identify the maximum absolute binary mixing enthalpy, denoted as $|\Delta H|$, among constituent binary pairs within the alloy. For this particular pair, we use ΔH multiplied by a factor as our feature. For example, for an alloy $A_aB_bC_c$, if $|\Delta H_{AB}|$ is the maximum absolute binary mixing enthalpy, ΔH_{max} is calculated as follows:

$$\Delta H_{\text{max}} = \frac{2 \times a \times b}{a + b} \times \Delta H_{AB}. \quad (6)$$

Here, ΔH_{AB} is obtained from the modified Miedema model [48]. The factor $\frac{2 \times a \times b}{a + b}$ accounts for the fractional number of A-B bonds in the alloy. The choice of “2” in $\frac{2 \times a \times b}{a + b}$ is a normalizing factor.

7. Mean heat of mixing ΔH_{mean} : ΔH_{mean} of the alloy represents the weighted average binary mixing enthalpy of all constituent binary pairs within the alloy. For a ternary alloy $A_aB_bC_c$, ΔH_{mean} is calculated as follows:

$$\Delta H_{\text{mean}} = \frac{(a+b) \times \Delta H_{\text{AB}} + (a+c) \times \Delta H_{\text{AC}} + (b+c) \times \Delta H_{\text{BC}}}{2 \times (a+b+c)} \quad (7)$$

2.3. Machine learning model

2.3.1. Random forest classification

In this study, we employ the random forest ML model to create classification models that map features (as described in Section 2.2) to the GFA of alloys. Random forest is robust to outliers, versatile, able to handle non-linear data and high dimensional data [49]. A random forest classification model builds numerous decision trees during training, and the model's prediction is determined by the label chosen by the majority of these decision trees. We use the open-source python package Scikit-learn to construct the random forest ML model. Grid search has been employed throughout the training process to optimize hyperparameters for the model, such as the number of decision trees, the number of features to choose from at each tree node, and the maximum depth of each tree, to obtain the highest classification accuracy. The trained model can assess the relative probability of an unknown alloy belonging to each GFA category, thus can categorize any new alloy into these predefined categories. Therefore, we can utilize the ML model to make predictions in the unknown composition space.

2.3.2. Within-group and out-of-group tests

To assess the effectiveness of Inoue's empirical rules and subgrouping for metallic glasses, we perform two sets of tests. In the first set of tests, we train the models on each subgroup and evaluate their performance on held-out test datasets within the same subgroup. We compare the prediction accuracy of the models trained on each subgroup with the models trained on the entire dataset. Such comparison allows us to quantify performance differences between the models and specifically determine whether physics-informed subgrouping enhances the prediction accuracy. To evaluate model performance, we use the classification accuracy as the metric, which measures the percentage of correctly classified observations by a given model.

In another complementary test of the effectiveness of subgrouping, we utilize the models trained on each subgroup to predict into other subgroups and compare these predictions with predictions made within the same subgroup (Fig. 2). To calibrate these results, we conduct a parallel experiment in which we randomly divide all data into seven subgroups, each of the same size as Inoue's seven subgroups. We then repeat the same process of ML training and prediction. Subsequently, we compare the results of this benchmark experiment using random subgrouping with those derived from Inoue's subgrouping.

3. Results and discussion

3.1. No subgrouping versus effective subgrouping: quantity versus quality

We found that the models trained on specific subgroups of data outperform the models trained on the entire dataset. Such out-performance suggests that physics-informed subgrouping enhances prediction accuracy. The average test accuracies are $96 \pm 2\%$ for models trained on subgroups of data, $81 \pm 7\%$ for models trained on all data, and $61 \pm 7\%$ for models trained on all data excluding data from the same subgroup (Fig. 3).

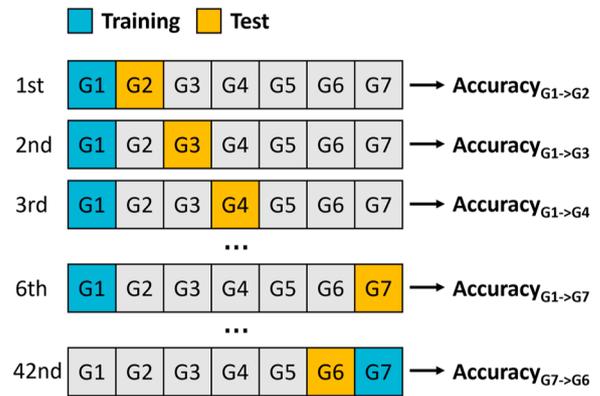


Fig. 2. Schematic of out-of-group testing. In the second set of tests proposed in this study to evaluate Inoue's subgrouping, the ML model is trained using data exclusively from one subgroup, e.g., G1. The trained model is then used to predict into other subgroups (G2, G3, etc.) allowing for a comparison of out-of-group prediction accuracies with those obtained from predictions within the same subgroup. Additionally, the testing procedure is repeated for a benchmark experiment where the data is randomly grouped into seven subgroups, following the same methodology described above.

3.2. Within-group tests versus out-of-group tests

The previous results on the outperformance of the models based on subgroup predictions compared to the model trained on all data reveal the effectiveness of subgrouping. Further and complimentary evidence for the effectiveness of such subgrouping can be seen when comparing within-group tests and out-of-group tests. Specifically, high within-group prediction and low out-of-group prediction would be indicative of an effective subgrouping. As shown in Fig. 4, within subgroups, the average prediction accuracy is $97 \pm 1\%$ (diagonal), which is significantly higher than the average accuracy of predictions into other groups, $55 \pm 3\%$ (off diagonal). The ratio of out-of-group accuracy to within-group accuracy is 0.57.

In the benchmark experiment (Fig. 4b), where we randomly group all data into seven "subgroups" of the same size as Inoue's seven subgroups, we found no significant differences between within-group predictions and out-of-group predictions, with an average accuracy of $88 \pm 3\%$ and $86 \pm 1\%$, respectively. The average accuracy of within-group predictions using random subgrouping is lower than that using Inoue's grouping, i.e., 88 versus 96%. The ratio of out-of-group accuracy to within-group accuracy is close to unity, 0.98. This ratio, which we define as Subgroup Similarity Score (SSS) indicates the specificity and meaningfulness of the grouping strategy. A lower SSS suggests that the predictions within a subgroup are significantly more accurate than predictions into other subgroups. This indicates that the subgrouping strategy effectively separates distinct subgroups, making it more meaningful for modeling the specific mechanisms within each subgroup. On the other hand, a higher SSS implies that there is higher similarity between the subgroups, making it challenging to distinguish between them. In such cases, the subgrouping strategy may not effectively capture the variations of underlying physics in material behavior, potentially limiting its usefulness in ML modeling. Values of SSS can range from 0 to 1 (approximately), with 0 representing the highest level of meaningfulness and distinction whereas 1 indicates the absence of any meaningful subgroups. Our results imply that Inoue's grouping strategy has significant physical meaning.

3.3. Feature importance: different gfa mechanisms in different subgroups

To further investigate the origin of effective subgrouping, we identify and analyze feature importance in our random forest ML models trained on different subgroups. Feature importance is a measure of the relative

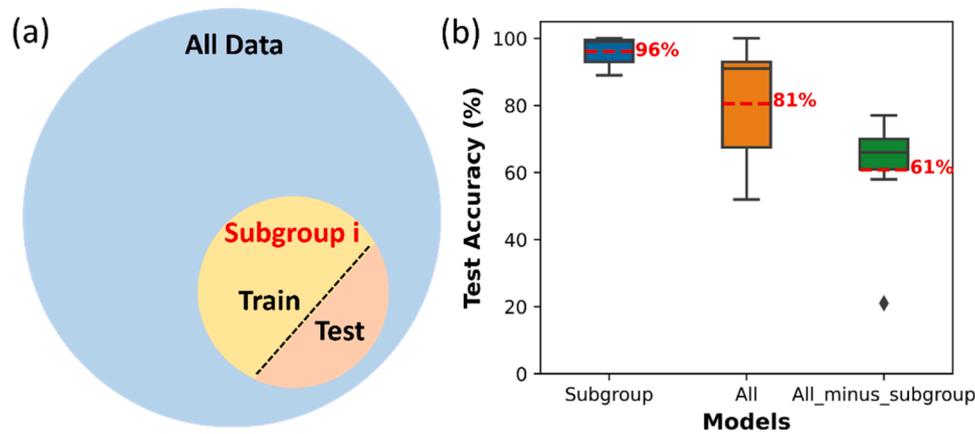


Fig. 3. Comparison of ML model performance between models trained on all data and models trained on Inoue's subgroups. (a) Visualization of different data partitions. (b) The average test accuracies are $96 \pm 2\%$ for models trained on subgroups of data, $81 \pm 7\%$ for models trained on all data, and $61 \pm 7\%$ for models trained on all data excluding data from the same subgroup. The error originates from the variance of prediction accuracies across seven subgroups.

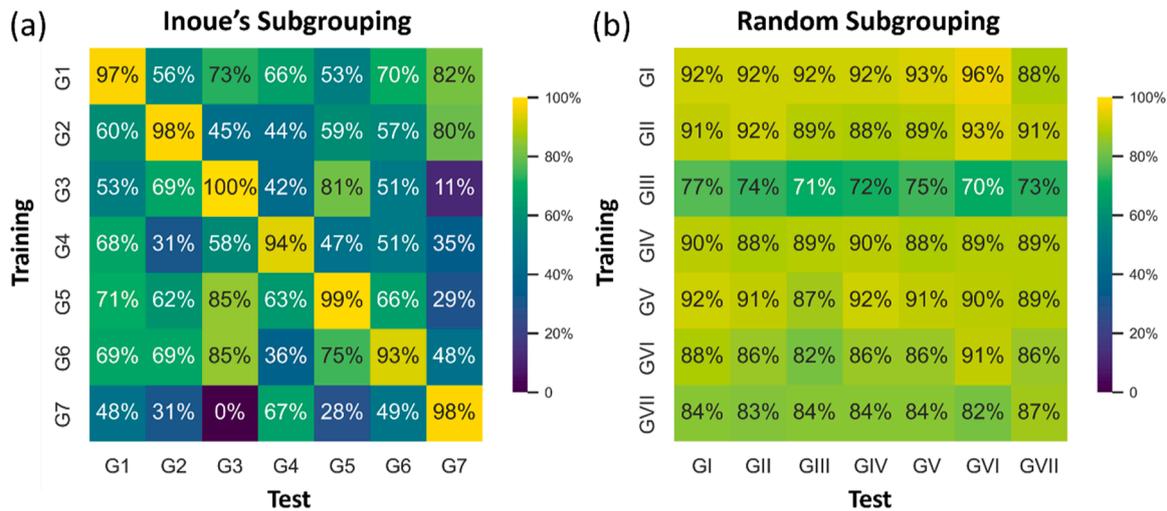


Fig. 4. Comparison of ML prediction accuracies between within-group predictions and out-of-group predictions using Inoue's subgrouping and random subgrouping. (a) For Inoue's subgrouping the average prediction accuracy of within-group tests is $97 \pm 1\%$ (diagonal), while the average accuracy of out-of-group tests is $55 \pm 3\%$ (off diagonal). The Subgroup Similarity Score (SSS), defined as the ratio of out-of-group accuracy to within-group accuracy, is 0.57. (b) For random subgrouping, the average prediction accuracy of within-group tests and out-of-group tests are $88 \pm 3\%$ and $86 \pm 1\%$, respectively, with a SSS of 0.98.

importance of each feature in making accurate predictions. We found that feature importance varies across groups, indicating distinct and subgroup-specific motifs for glass formation (Fig. 5). Notably, each subgroup exhibited a different most important feature: T_r for G1, Δr for G2, ϕ for G4 and G5, δ for G6, and ΔH_{mean} for G7.

For example, when examining alloys in G6 (ETM/Ln-LTM/BM), it is difficult to find sufficiently large atomic size difference, especially among the predominant transition metals. Consequently, atomic size difference (δ) should be expected to be a dominating factor in determining the GFA within this group and this is indeed the feature with the highest feature importance in our analysis. Other features, such as heat of mixing (ΔH_{max} and ΔH_{mean}) and liquidus temperature reduction (T_r) are widely available to lead to high GFA, hence do not play such a dominant role.

On the other hand, for alloys in G7 (IIA-LTM/BM), atomic size range and ratio are typically large, while the heat of mixing among elements is relatively small and similar. Hence, heat of mixing (ΔH_{mean}) is the dominating feature, whereas atomic size range (Δr) and ratio (ϕ) are much less important. Again, this has also been found in our feature importance analysis.

In summary, the feature importance analysis supports the

effectiveness of Inoue's subgrouping. Further the analysis aligns well with the fact that a variety of different attributes of an alloy can contribute to glass formation, supporting the diverse and rich nature of this phenomenon.

4. Discussion

The results depicted in Fig. 3 clearly demonstrate that the models trained on individual subgroups consistently outperform the model trained on the entire dataset. The observed improvement in prediction accuracy reflects their large range of constitutive elements and their associated attributes in absolute and relative-to-each-other terms, which suppresses crystallization in different ways [41,45,46,50-59]. By partitioning the dataset into subgroups based on Inoue's subgrouping strategies, we effectively capture the distinct characteristics and underlying physics associated with each subgroup. This allows the individual models, despite using fewer data for training, to better learn and leverage the specific patterns and behaviors exhibited within each alloy subgroup. Consequently, when presented with new observations from the same subgroup during testing, the models can make more accurate predictions. By isolating into subgroups, specific strategies to form

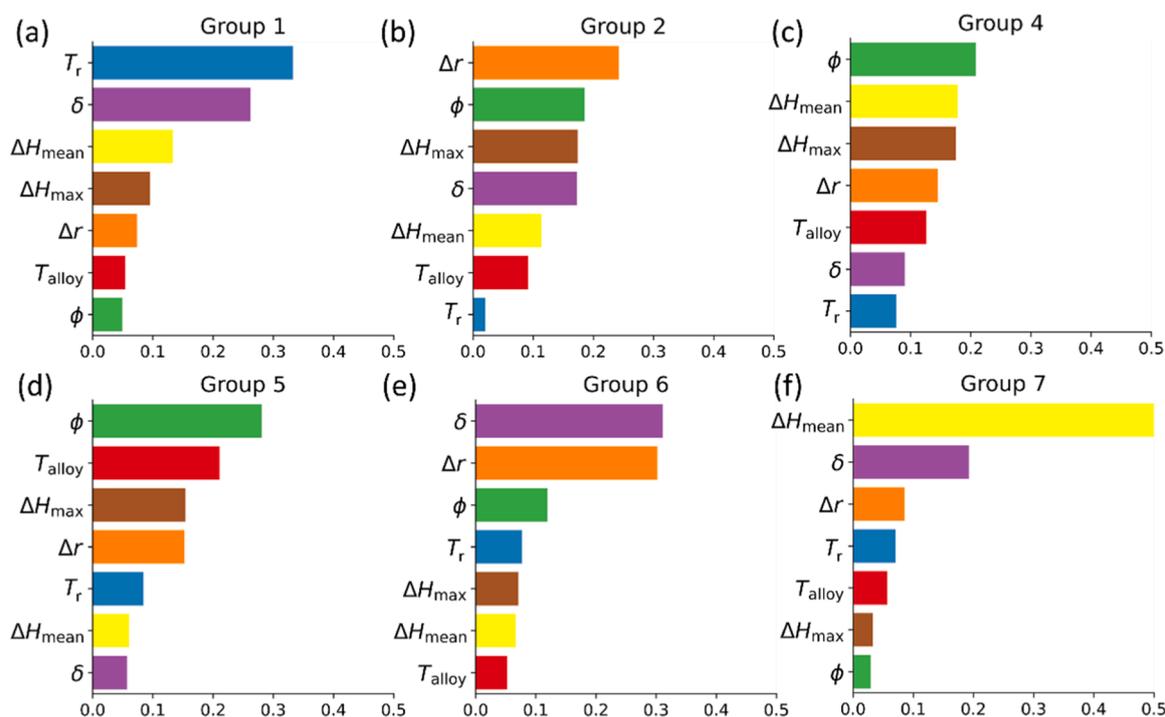


Fig. 5. Variation in feature importance of ML models suggests different GFA mechanisms are at play in different subgroups. The features are: T_{alloy} , liquidus temperature of the alloy; T_r , liquidus temperature reduction; δ , atomic size difference; ϕ , atomic size ratio; Δr , atomic size range; ΔH_{max} , maximum heat of mixing; ΔH_{mean} , mean heat of mixing. G3 exists solely of non-glass formers, hence does not have a feature importance plot.

glasses can be revealed that would otherwise be smeared out when considering the entire group of metallic glasses. Such subgrouping-based models specializing in modeling the specific mechanisms of each subgroup can be expected to yield higher prediction accuracy.

To explore why predictions into other subgroups are limited, we calculate the confusion matrix for out-of-group predictions (Table 1). The confusion matrix goes beyond simple average accuracies and reveals specific alloys that result in false positive or false negative classifications. Thus, we can gain insights into how well models perform within specific subgroups and identify any patterns of misclassification.

We first analyzed the characteristics of Inoue's seven subgroups: G1, G5, and G7 BMGs exhibit a tendency where the main alloying element possesses the largest atomic radius among all alloying elements and forms an atomic pair with the largest negative value of heat of mixing. BMGs belonging to G2 and G4 have a main element with an intermediate atomic radius, independent of the atomic pair with the largest negative heat of mixing. For multicomponent BMGs in G3, the main element can either be the larger-sized element in the atomic pair with the largest negative heat of mixing or an element within the same group in the periodic table. BMGs in G6 are characterized by the main alloying element with the smallest atomic radius among the other alloying elements, while still being an element in the atomic pair with the largest negative heat of mixing.

Using the confusion matrix results allow us to conclude for example that the presence of metalloid elements gives rise to different mechanisms for glass formation. When we use the ML model trained on G1 (ETM/Ln-LTM/BM-Al/Ga) to predict into G2 (ETM/Ln-LTM/BM-Metalloid), we observe a high false positive rate (FPR) of 73 % (false positives refer to non-glass formers wrongly classified as BMGs) and a false negative rate (FNR) of 4 %, indicating an overestimation of GFA for alloys containing metalloids in G2. This can be explained by the fact that metalloid elements have a significantly smaller atomic size compared to other elements, resulting in larger differences in atomic size difference, ratio, and range within the alloy. The model trained on G1 has primarily encountered a limited range (on the smaller end of the spectrum) of atomic size difference, ratio, and range during training, on which the

decision rules for GFA were based on. Consequently, the model overestimates GFA in G2, as the values of these features in G2 are likely to exceed the thresholds established in the G1 model.

A similar case arises when using the G4 (IIA-ETM/Ln-LTM/BM) model to predict into G2 (ETM/Ln-LTM/BM-Metalloid) where atomic size differences are larger due to the existence of metalloid elements in G2. The confusion matrix reveals an FPR of 76 % and an FNR of 0 %, suggesting an overestimation of GFA. Analogous results and analysis can be found when applying the ML model trained on G7 (IIA-LTM/BM) to predict into G2 (ETM/Ln-LTM/BM-Metalloid).

The pairs of groups, G5 (LTM/BM-Metalloid) and G7 (IIA-LTM/BM), exhibit intriguing similarities in their characteristics. Both groups share the common trait of having the main alloying element with the greatest atomic radius among the other alloying elements. Additionally, this main element is part of the atomic pair with the largest negative heat of mixing. Training on G7 and testing on G5 resulted in an FPR of 80 % and an FNR of 0 %. Conversely, training on G5 and testing on G7 yielded an FPR of 0 % and an FNR of 89 %. These findings indicate a consistent overestimation in one direction (from G7 to G5) and a contrasting underestimation in the reverse direction. Notably, the main element in G7 is Ca, which possesses one of the largest atomic radii of 0.197 nm. The presence of Ca in all of G7 alloys diminishes the significance of the "atomic size ratio ϕ " feature, while emphasizing other features related to heat of mixing, as depicted in Fig. 5. In contrast, G5 comprises metalloid elements such as B, C, P, and Si, which have smaller atomic radii ranging from 0.077 to 0.117 nm. Additionally, G5 alloys exhibit relatively higher heat of mixing values ranging from -55 to -34.5 kJ/mol for combinations between the major elements (Fe, Ni, Co, and Pd) and the metalloids (B, C, P, and Si) in the LTM/BM. These distinct characteristics and the inherent large negative mixing in most G5 alloys reduce the influence of heat of mixing-related features when estimating other groups. Consequently, the high importance of the 'atomic size ratio ϕ ' feature leads to an overestimation of GFA for G5 when the model is trained using G7, while underestimating the GFA of G7 when trained on G5. Similar patterns of reversed FPR/FNR ratios, such as G4-G5, and G2-G6, can be explained by employing a similar logic to the G5-G7 pair.

Table 1

Confusion matrix for out-of-group predictions using Inoue's subgrouping. The table illustrates a confusion matrix summarizing the results of GFA prediction in out-of-group tests. The matrix provides a breakdown of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), presented in percentages, where "positive" indicates BMG, and "negative" denotes non-glass formers. Additionally, it provides key metrics such as the false positive rate (FPR), false negative rate (FNR), and overall accuracy (ACC).

Train	Test	Total #	TP %	TN %	FP %	FN %	FPR %	FNR %	ACC %
G1	G2	394	30	19	50	1	73	4	49
G1	G3	33	0	70	30	0	30	n/a	70
G1	G4	264	52	14	28	6	67	10	66
G1	G5	355	9	44	37	10	45	54	53
G1	G6	191	28	39	12	21	23	43	67
G1	G7	119	87	1	10	3	92	3	87
G2	G1	724	38	20	32	10	62	20	58
G2	G3	33	0	42	58	0	58	n/a	42
G2	G4	264	26	16	26	32	62	55	42
G2	G5	355	10	50	32	9	39	49	59
G2	G6	191	43	11	40	5	79	11	54
G2	G7	119	76	7	4	13	38	14	83
G4	G1	724	39	28	24	8	46	17	67
G4	G2	394	31	16	52	0	76	0	48
G4	G3	33	0	85	15	0	15	n/a	85
G4	G5	355	18	41	40	1	49	3	60
G4	G6	191	16	41	10	32	19	67	58
G4	G7	119	32	10	1	57	8	64	42
G5	G1	724	24	47	6	23	11	49	71
G5	G2	394	1	61	8	30	11	96	62
G5	G3	33	0	85	15	0	15	n/a	85
G5	G4	264	26	39	3	32	8	55	65
G5	G6	191	29	41	10	19	19	40	71
G5	G7	119	18	11	0	71	0	80	29
G6	G1	724	19	50	2	28	4	59	70
G6	G2	394	1	69	0	31	0	98	69
G6	G3	33	0	85	15	0	15	n/a	85
G6	G4	264	10	32	10	47	24	82	42
G6	G5	355	2	74	7	17	8	88	77
G6	G7	119	38	6	5	51	46	58	44
G7	G1	724	47	1	52	0	98	0	48
G7	G2	394	31	0	69	0	100	0	31
G7	G3	33	0	0	100	0	100	n/a	0
G7	G4	264	55	13	30	2	71	4	68
G7	G5	355	19	9	72	0	89	0	28
G7	G6	191	47	3	49	2	95	4	49

The insights gained from the pairs found in the confusion matrix, which demonstrate a consistent overestimation in one direction and underestimation in the other direction, highlight the critical importance of subgrouping based on the underlying physics. This approach proves crucial, not only for accurately predicting the GFA of specific alloy groups but also for understanding the role of each individual feature. By incorporating the relevant physical characteristics and properties of the alloys into the subgrouping and analyzing them using machine learning techniques, we can achieve more robust predictions and gain a deeper understanding of the underlying mechanisms governing the scientific problem at hand. This integration of physics-based subgrouping and machine learning methods offers a powerful framework for advancing our knowledge in materials science and facilitating the design and development of novel materials with tailored properties.

The above discussion revealed the need and power of physical insights (physics-informed subgrouping and features). Obviously, the question arises whether there are ML or data science strategies that reveal such subgrouping or features without using physical insights. In other words, are physical insights necessary for building accurate ML models? To test whether subgrouping (and hence ML models based upon such groups) without physical insights is at least as useful as physical insights, we employ clustering or subgroup discovery strategies that are based solely upon the characteristics of the features of the data. Note that while these features (described in Section 2.2) are derived from physical principles, we will use only these features to group the alloys

and will not consider whether such features result in the formation of glass or not.

One widely used strategy to group data into subgroups is k -means clustering in which k , the number of averages, is specified and then data points are grouped into k clusters or subgroups according to the distances of their features from these average values. In the next computational experiment, we show that simply using the features alone (without additional physical insights) with the k -means clustering algorithm is not as effective in building predictive ML models as the physically derived groups, although such clustering is better than random grouping.

We run k -means clustering for $k = 7$ on the data (Section 2.1) and generate seven subgroups. We perform the same within-group and out-of-group prediction experiments (as described in Section 2.3.2) using the k -means clusters. Fig. 6 shows the within-group predictions and the out-of-group predictions. We find that the within-group predictions are worse for the k -means cluster than for Inoue's groups (especially on G2, G3, and G5), albeit a bit better on G1, G4, and G6. Interestingly, the out-of-group predictions are considerably higher for the k -means clusters than for Inoue's groups, suggesting that this subgrouping does not separate or distinguish amongst the subgroups nearly as well. Specifically, the average prediction accuracy of within-group tests is $96 \pm 1\%$ and that of out-of-group tests is $63 \pm 3\%$, resulting in a SSS of 0.66, higher than Inoue's grouping with a SSS of 0.57. The k -means subgrouping is reasonably effective at prediction but not effective at revealing underlying structure nor as useful for revealing physical mechanisms that might drive glass formation for different types of alloys.

The k -means algorithm clusters data based on their features alone, without taking into account any labels. Obviously, when labels are not considered, no conclusions can be drawn about the feature importance and, thus, k -means clusters do not reflect the contribution of features. In cases where there are different underlying mechanisms and physics defining the GFA, clustering analysis alone cannot be efficient in identifying subgroups that would be statistically correlated with different labels. Therefore, the effectiveness of clustering analysis that does not make use of any other information depends on whether the data has the same feature importance and underlying mechanisms for the specific problem being studied.

We argue that to generate a physically meaningful clustering of alloys into subgroups, the labels (e.g., if a certain feature vector is

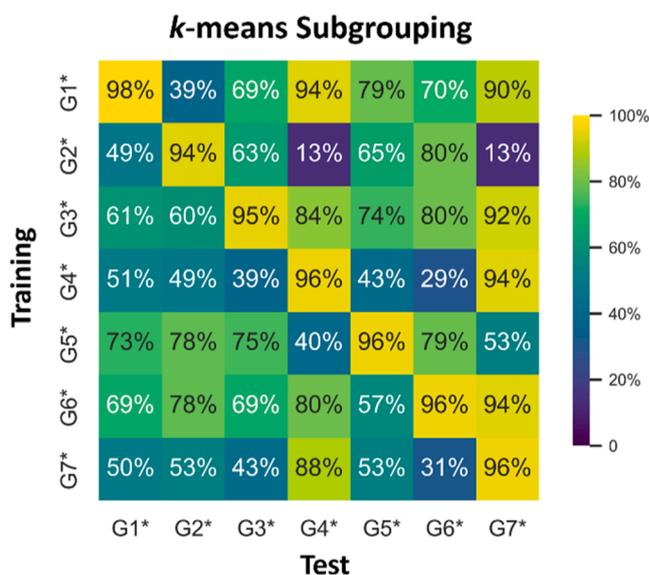


Fig. 6. ML prediction accuracies between within-group predictions and out-of-group predictions using k -means clustering subgrouping strategy. The average prediction accuracy of within-group tests is $96 \pm 1\%$ and the average prediction accuracy of out-of-group tests is $63 \pm 3\%$, resulting in a SSS of 0.66.

associated with a glass or not) must be considered. In theory, the best subgrouping could be determined through an optimization process. This optimization process would find the m disjoint, non-empty groupings for the n data points by maximizing the within-group accuracy and minimizing the out-of-group accuracy. The search space of all possible groupings is exponentially large (it is of size roughly $n^m/(m!)$ for each possible value of m [60]). For 1000 alloys, the number of groupings of size 7 is approximately 2×10^{17} , already far too large for us to optimize over. Indeed, all subgroup discovery algorithms define some form of objective function (i.e., quantity that one wishes to optimize) and then restricts in some *ad hoc* fashion the size of the search space (from exponential to hopefully a polynomial-sized space in n and m to generate an algorithm which runs in a reasonable amount of time). This is the idea behind all subgroup algorithms [61–64]. We emphasize that all such approximations or restrictions are *ad hoc* and data-driven, rather than physically motivated.

As a result of the k -means experiment and the discussion above, we seek a middle ground or a compromise between the solely data-driven clustering (or more generally, the *ad hoc* restriction of the subgroupings) and the ‘hand crafted’ groupings of Inoue. We advocate using both algorithmic heuristics plus physical principles or insights to guide the development of effective ML models in material science.

5. Conclusion

In conclusion, this study underscores the pivotal role of effective subgrouping in enhancing ML predictions for complex materials science phenomena, particularly in the context of metallic glass formation. By applying Inoue’s subgrouping approach, which are based on physical insights, we have demonstrated that such subgrouping significantly improves prediction accuracy. Moreover, our results highlight the potential of ML and data science strategies to quantitatively assess guiding principles utilized in materials science, allowing for a rigorous evaluation of empirical rules like Inoue’s. Additionally, our investigation unveils the existence of diverse mechanisms and contributions controlling GFA within these subgroups. These subgroup-specific behaviors emphasize the importance of subgrouping based on underlying physics and chemical attributes, as it allows for a more accurate modeling of the specific mechanisms governing BMG formation in each subgroup.

This study advocates a synergistic approach that combines algorithmic heuristics with physical insights to advance our understanding of complex materials science and improve the predictive capabilities of ML models. Such an integrated framework has the potential to unlock new insights into material composition-property relationships and accelerate materials discovery in a wide range of applications beyond metallic glass formation.

Author contributions

J.S. and G.L. conceived the work. G.L. performed the machine learning study and analyzed the results. G.L. and J.S. wrote the manuscript with input from S.S., A.C.G., and C.S.O. All authors contributed to the discussion and proofread of the manuscript.

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.

Acknowledgment

This work was financially supported by the National Science Foundation through NSF-DMR (grant no. 2104316).

References

- [1] A. Samanta, M.E. Tuckerman, T.Q. Yu, E. Wn, Microscopic mechanisms of equilibrium melting of a solid, *Science* 346 (6210) (2014) 729–732.
- [2] G.L.W. Hart, T. Mueller, C. Toher, S. Curtarolo, Machine learning for alloys, *Nat. Rev. Mater.* 6 (8) (2021) 730–755.
- [3] G. Deffrennes, K. Terayama, T. Abe, E. Ogamino, R. Tamura, A framework to predict binary liquidus by combining machine learning and CALPHAD assessments, *Mater. Des.* 232 (2023) 112111.
- [4] S.A. Kube, S. Sohn, R. Ojeda-Mota, T. Evers, W. Polsky, N. Liu, K. Ryan, S. Rinehart, Y. Sun, J. Schroers, Compositional dependence of the fragility in metallic glass forming liquids, *Nat. Commun.* 13 (1) (2022) 3708.
- [5] N. Liu, S. Sohn, M.Y. Na, G.H. Park, A. Raj, G. Liu, S.A. Kube, F. Yuan, Y. Liu, H. J. Chang, J. Schroers, Size-dependent deformation behavior in nanosized amorphous metals suggesting transition from collective to individual atomic transport, *Nat. Commun.* 14 (1) (2023) 5987.
- [6] 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 (6366) (2017) 1033–1037.
- [7] W.J. Boettinger, S.R. Coriell, A.L. Greer, A. Karma, W. Kurz, M. Rappaz, R. Trivedi, Solidification microstructures: recent developments, future directions, *Acta Mater.* 48 (1) (2000) 43–70.
- [8] K. Tsutsui, H. Terasaki, K. Uto, T. Maemura, S. Hiramatsu, K. Hayashi, K. Moriguchi, S. Morito, A methodology of steel microstructure recognition using SEM images by machine learning based on textural analysis, *Mater. Today Commun.* 25 (2020).
- [9] R.Q. Liu, A. Kumar, Z.Z. Chen, A. Agrawal, V. Sundararaghavan, A. Choudhary, A predictive machine learning approach for microstructure optimization and materials design, *Sci. Rep.* 5 (2015).
- [10] J.N. Fuhg, L. van Wees, M. Obstalecki, P. Shade, N. Bouklas, M. Kasemer, Machine-learning convex and texture-dependent macroscopic yield from crystal plasticity simulations, *Materialia* 23 (2022).
- [11] G.N. Liu, S. Sohn, N.J. Liu, A. Raj, U.D. Schwarz, J. Schroers, Single-Crystal Nanostructure Arrays Forming Epitaxially through Thermomechanical Nanomolding, *Nano Lett.* 21 (23) (2021) 10054–10061.
- [12] J. Schroers, Bulk metallic glasses, *Phys. Today* 66 (2) (2013) 32.
- [13] A. Inoue, Stabilization of metallic supercooled liquid and bulk amorphous alloys, *Acta Mater.* 48 (1) (2000) 279–306.
- [14] G.N. Liu, S. Sohn, S.A. Kube, A. Raj, A. Mertz, A. Nawano, A. Gilbert, M. D. Shattuck, C.S. O’Hern, J. Schroers, Machine learning versus human learning in predicting glass-forming ability of metallic glasses, *Acta Mater.* 243 (2023).
- [15] N.J. Liu, T.X. Ma, C.Q. Liao, G.N. Liu, R.M.O. Mota, J.B. Liu, S. Sohn, S. Kube, S. F. Zhao, J.P. Singer, J. Schroers, Combinatorial measurement of critical cooling rates in aluminum-base metallic glass forming alloys, *Sci. Rep.* 11 (1) (2021).
- [16] W.L. Johnson, Bulk glass-forming metallic alloys: science and technology, *MRS Bull.* 24 (10) (1999) 42–56.
- [17] A.L. Greer, M.B. Costa, O.S. Houghton, Metallic glasses, *MRS Bull.* (2023).
- [18] J. Schroers, W.L. Johnson, Ductile bulk metallic glass, *Phys. Rev. Lett.* 93 (25) (2004).
- [19] W.L. Johnson, K. Samwer, A universal criterion for plastic yielding of metallic glasses with a $(T/T_g)^{2/3}$ temperature dependence, *Phys. Rev. Lett.* 95 (19) (2005).
- [20] L. Shao, J. Ketkaew, P. Gong, S.F. Zhao, S. Sohn, P. Bordeenithikasem, A. Datye, R. M.O. Mota, N.J. Liu, S.A. Kube, Y.H. Liu, W. Chen, K.F. Yao, S.J. Wu, J. Schroers, Effect of chemical composition on the fracture toughness of bulk metallic glasses, *Materialia* 12 (2020).
- [21] M.F. Ashby, A.L. Greer, Metallic glasses as structural materials, *Scr. Mater.* 54 (3) (2006) 321–326.
- [22] S. Sohn, N. Liu, G.H. Yoo, A. Ochiai, J. Chen, C. Levitt, G. Liu, S.C. Schroers, E. T. Lund, E.S. Park, J. Schroers, A framework for plasticity in metallic glasses, *Materialia* 31 (2023) 101876.
- [23] W. Chen, Z. Liu, J. Ketkaew, R.M.O. Mota, S.H. Kim, M. Power, W. Samela, J. Schroers, Flaw tolerance of metallic glasses, *Acta Mater.* 107 (2016) 220–228.
- [24] J.J. Kruzic, Bulk metallic glasses as structural materials: a review, *Adv. Eng. Mater.* 18 (8) (2016) 1308–1331.
- [25] L.M. Lai, R.G. He, K.L. Ding, T.H. Liu, R.Y. Liu, Y.X. Chen, S.F. Guo, Ternary Co-Mo-B bulk metallic glasses with ultrahigh strength and good ductility, *J. NonCryst. Solids* 524 (2019).
- [26] M. Zhang, Y.J. Wang, G.Z. Ding, R. Zheng, R.F. Wei, G.Y. Zhang, Q.J. Sun, X. J. Zhao, L. Liu, Biocompatible superhydrophobic surface on Zr-based bulk metallic glass: fabrication, characterization, and biocompatibility investigations, *Ceram. Int.* 49 (15) (2023) 25549–25562.
- [27] J. Schroers, Processing of bulk metallic glass, *Adv. Mater.* 22 (14) (2010) 1566–1597.
- [28] R. Busch, J. Schroers, W.H. Wang, Thermodynamics and kinetics of bulk metallic glass, *MRS Bull.* 32 (8) (2007) 620–623.
- [29] T. Egami, C.W. Ryu, Why Is the Range of Timescale So Wide in Glass-Forming Liquid? *Front. Chem.* 8 (2020).
- [30] M.E. Blodgett, T. Egami, Z. Nussinov, K.F. Kelton, Proposal for universality in the viscosity of metallic liquids, *Sci. Rep.* 5 (2015).
- [31] T. Egami, Atomic level stresses, *Prog. Mater. Sci.* 56 (6) (2011) 637–653.

- [32] G.R. Garrett, M.D. Demetriou, M.E. Launey, W.L. Johnson, Origin of embrittlement in metallic glasses, *P. Natl. Acad. Sci. USA* 113 (37) (2016) 10257–10262.
- [33] W. Hume-Rothery, H.M. Powell, On the theory of super-lattice structures in alloys, *Z. Kristallogr.* 91 (1) (1935) 23–47.
- [34] W. Hume-Rothery, R.E. Smallman, C.W. Haworth, *The Structure of Metals and Alloys*, Metals & Metallurgy Trust, London, 1969, 5th ed.
- [35] E. Hall, The deformation and ageing of mild steel: III discussion of results, *Proc. Phys. Soc. Lond. Sect. B* 64 (9) (1951) 747.
- [36] N. Petch, The cleavage strength of polycrystals, *J. Iron Steel Inst.* 174 (1953) 25–28.
- [37] D. Turnbull, Under what conditions can a glass be formed, *Contemp. Phys.* 10 (5) (1969) 473–&.
- [38] J.J. Lewandowski, W.H. Wang, A.L. Greer, Intrinsic plasticity or brittleness of metallic glasses, *Philos. Mag. Lett.* 85 (2) (2005) 77–87.
- [39] T. Chookajorn, H.A. Murdoch, C.A. Schuh, Design of stable nanocrystalline alloys, *Science* 337 (6097) (2012) 951–954.
- [40] W.L. Johnson, J.H. Na, M.D. Demetriou, Quantifying the origin of metallic glass formation, *Nat. Commun.* 7 (2016).
- [41] K. Zhang, Y.H. Liu, J. Schroers, M.D. Shattuck, C.S. O'Hern, The glass-forming ability of model metal-metalloid alloys, *J. Chem. Phys.* 142 (10) (2015).
- [42] J. Orava, A.L. Greer, Fast and slow crystal growth kinetics in glass-forming melts, *J. Chem. Phys.* 140 (21) (2014).
- [43] S.A. Kube, S. Sohn, D. Uhl, A. Datye, A. Mehta, J. Schroers, Phase selection motifs in High Entropy Alloys revealed through combinatorial methods: large atomic size difference favors BCC over FCC, *Acta Mater.* 166 (2019) 677–686.
- [44] Y. Zhang, Y.J. Zhou, J.P. Lin, G.L. Chen, P.K. Liaw, Solid-solution phase formation rules for multi-component alloys, *Adv. Eng. Mater.* 10 (6) (2008) 534–538.
- [45] A. Takeuchi, A. Inoue, Classification of bulk metallic glasses by atomic size difference, heat of mixing and period of constituent elements and its application to characterization of the main alloying element, *Mater. Trans.* 46 (12) (2005) 2817–2829.
- [46] Y.L. Li, S.F. Zhao, Y.H. Liu, P. Gong, J. Schroers, How many bulk metallic glasses are there? *ACS Comb. Sci.* 19 (11) (2017) 687–693.
- [47] Y. Kawazoe, Nonequilibrium phase diagrams of ternary amorphous alloys, *LB: New Ser., Group III: Condens.* 37 (1997) 1–295.
- [48] D.B. Miracle, D.V. Louzguine-Luzgin, L.V. Louzguina-Luzgina, A. Inoue, An assessment of binary metallic glasses: correlations between structure, glass forming ability and stability, *Int. Mater. Rev.* 55 (4) (2010) 218–256.
- [49] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [50] T. Egami, Y. Waseda, Atomic size effect on the formability of metallic glasses, *J. NonCryst. Solid.* 64 (1–2) (1984) 113–134.
- [51] A. Takeuchi, A. Inoue, Analyses of characteristics of atomic pairs in ferrous bulk metallic glasses using classification of bulk metallic glasses and pettifor map, *J. Optoelectron. Adv. Mater.* 8 (5) (2006) 1679–1684.
- [52] A.L. Greer, Materials science - confusion by design, *Nature* 366 (6453) (1993) 303–304.
- [53] J.Q. Wang, N. Chen, P. Liu, Z. Wang, D.V. Louzguine-Luzgin, M.W. Chen, J. H. Perepezko, The ultrastable kinetic behavior of an Au-based nanoglass, *Acta Mater.* 79 (2014) 30–36.
- [54] D.B. Miracle, A structural model for metallic glasses, *Nat. Mater.* 3 (10) (2004) 697–702.
- [55] E.S. Park, H.J. Chang, D.H. Kim, Improvement of glass-forming ability and phase separation in Cu-Ti-rich Cu-Ti-Zr-Ni-Si bulk metallic glasses, *J Alloy Compd* 504 (2010) S27–S30.
- [56] H. Assadi, J. Schroers, Crystal nucleation in deeply undercooled melts of bulk metallic glass forming systems, *Acta Mater.* 50 (1) (2002) 89–100.
- [57] Y.C. Hu, J. Schroers, M.D. Shattuck, C.S. O'Hern, Tuning the glass-forming ability of metallic glasses through energetic frustration, *Phys. Rev. Mater.* 3 (8) (2019).
- [58] E. Perim, D. Lee, Y.H. Liu, C. Toher, P. Gong, Y.L. Li, W.N. Simmons, O. Levy, J. J. Vlassak, J. Schroers, S. Curtarolo, Spectral descriptors for bulk metallic glasses based on the thermodynamics of competing crystalline phases, *Nat. Commun.* 7 (2016).
- [59] J.H. Na, M.D. Demetriou, M. Floyd, A. Hoff, G.R. Garrett, W.L. Johnson, Compositional landscape for glass formation in metal alloys, *P. Natl. Acad. Sci. USA* 111 (25) (2014) 9031–9036.
- [60] R.P. Stanley, *Enumerative Combinatorics*, 1, *Cam St Ad M* 49, 2012, pp. 1–626, 2nd Edition.
- [61] B.R. Goldsmith, M. Boley, J. Vreeken, M. Scheffler, L.M. Ghiringhelli, Uncovering structure-property relationships of materials by subgroup discovery, *New J. Phys.* 19 (1) (2017).
- [62] C. Sutton, M. Boley, L.M. Ghiringhelli, M. Rupp, J. Vreeken, M. Scheffler, Identifying domains of applicability of machine learning models for materials science, *Nat. Commun.* 11 (1) (2020).
- [63] J.H. Friedman, N.I. Fisher, Bump hunting in high-dimensional data, *Stat. Comput.* 9 (2) (1999) 123–143.
- [64] S. Wrobel, An algorithm for multi-relational discovery of subgroups, *Lect. Note. Artif. Int.* 1263 (1997) 78–87.