

## Efficient learning strategy for predicting glass forming ability in imbalanced datasets of bulk metallic glasses

Xuhe Gong,<sup>1,2</sup> Jiazi Bi,<sup>1</sup> Xiaobin Liu,<sup>1</sup> Ran Li,<sup>1,\*</sup> Ruijuan Xiao<sup>2,†</sup>, Tao Zhang,<sup>1</sup> and Hong Li<sup>2</sup>

<sup>1</sup>*School of Materials Science and Engineering, Key Laboratory of Aerospace Materials and Performance (Ministry of Education), Beihang University, Beijing 100191, China*

<sup>2</sup>*Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*



(Received 16 December 2023; accepted 22 April 2024; published 10 May 2024)

The prediction of glass forming ability (GFA) and various properties in bulk metallic glasses (BMGs) pose a challenge due to the unique disordered atomic structure in this type of material. Machine learning shows the potential ability to find a way out. However, the training set from the experimental data of BMGs faces the issue of data imbalance, including the distribution of data related to elements, the range of performance data, and the distribution of sparse and dense data area in each specific system. In this work, the origin of the data imbalance and its impact on the GFA prediction ability of machine-learning models are analyzed. We propose the solutions by training the model using the pruned dataset to mitigate the imbalance and by performing an active experimental iterative learning to compensate for the information loss resulting from data reduction. The strategy is proved in Zr-Al-Cu system, and the automated workflow has been established. It effectively avoids the prediction results from trapping into the intensive training-data area or from inducing by the data distribution of similar element systems. This approach will expedite the development of BMGs compositions, especially for unexplored systems.

DOI: [10.1103/PhysRevMaterials.8.055602](https://doi.org/10.1103/PhysRevMaterials.8.055602)

### I. INTRODUCTION

In the era of big data, machine learning has gained tremendous popularity and adoption across various fields, including but not limited to medicine [1], wireless communication [2], and materials science [3–6]. Researchers often use machine learning to analyze and summarize the patterns in large-scale data due to its exceptional processing speed and data-mining capabilities. In the realm of materials science, machine learning has been extensively employed by researchers to facilitate property prediction, composition design, and mechanism investigation [7–10]. Especially for materials with demanding experimental preparation processes, high costs, and intricate influencing factors, machine learning plays a crucial role in expediting research progress, facilitating data analysis, and more [11,12].

Due to the unique disordered structure, metallic glasses (MGs) have various excellent properties such as high strength and hardness [13], excellent soft magnetic properties [14], and distinct catalytic activity [15], compared with crystalline materials. However, commercial application of MGs is limited by the low glass forming ability (GFA). The development of the MG composition with sufficient GFA and desired properties becomes the key in this field. The GFA can be directly characterized by the critical cooling rate ( $R_c$ ) or the critical casting diameter ( $D_{max}$ ), and the latter is more frequently taken as the quantity to indicate the GFA because it is easier to

measure than  $R_c$ . Researchers have spent a great deal of time searching for BMGs with  $D_{max}$  larger than 1 mm, and many high-GFA systems have been developed, such as Pd-Cu-Ni-P [16], Mg-Zn-Ca [17], Ti-Zr-Ni-Be [18], etc. Nevertheless, this is still a tiny fraction of the large number of potential systems, and there are still a large number of unknown systems worth further investigation. At present, the composition exploration of most bulk metallic glasses (BMGs) primarily depends on “trial and error” method. The target composition is verified one by one by the method of rapid melt quenching, significantly diminishing the efficiency of composition discovery. High-throughput experiments utilizing techniques such as magnetron sputtering have been employed to improve the efficiency of experimental verification [19–21]. However, it is important to note that the information obtained from thin-film samples produced by magnetron sputtering cannot be directly extended to bulk samples, because the cooling rate in magnetron sputtering is on the order of  $10^{12}$  K/s, while traditional BMGs prepared through methods like copper mold casting have cooling rates in the range of  $10^3$  to  $10^8$  K/s [22]. Despite the possibility of obtaining amorphous thin-film samples through sputtering under the same composition conditions, crystallization may still occur during the process of preparing bulk samples due to limitations in cooling rate. In the pursuit of high GFA in BMGs, it becomes imperative to establish efficient composition search methods capable of reducing economic costs and time costs simultaneously. One of the traditional approaches employed by researchers relies on empirical principles such as the eutectic point criterion and Inoue’s three principles, or some empirical parameters composed of thermodynamic parameters such as

\*liran@buaa.edu.cn

†rjxiao@aphy.iphy.ac.cn

$T_g(T_g/T_m, T_g/T_l)$  [23],  $\gamma[T_x/(T_l + T_g)]$  [24], and  $\gamma_m[(2T_x - T_l)/T_g]$  [25], where  $T_g$  denotes glass-transition temperature,  $T_x$  denotes crystallization temperature, and  $T_l$  denotes liquidus temperature. Although these parameters have been validated in certain systems, their applicability is limited when developing new materials. On one hand, in order to obtain characteristic temperature parameters such as  $T_g$ , it is necessary to conduct differential thermal analysis or differential scanning calorimetry experiments, which require the preparation of amorphous sample; thus, these descriptors cannot be obtained during the model prediction stage. On the other hand, the universality of such empirical parameters is also difficult to confirm [26]. It is of great significance to explore a more reliable method for predicting the composition-related properties of MGs, including but not confined to GFA, so as to facilitate quantitative composition design for BMGs.

On this basis, the machine-learning method has been adopted to study the different properties of MGs. You *et al.* [27] developed an artificial neural network model that aimed to classify the amorphous (crystalline) phases and correlate the excess electrical resistivity of the alloys to the full width at half maximum value of x-ray-diffraction (XRD) patterns. Ren *et al.* [28] successfully distinguished the composition regions capable of forming amorphous films in Co-Ti-Zr and Co-Fe-Zr systems through the iterative combination of high-throughput experiments and machine learning. Jeon *et al.* [29] established a random forest model to design Ni-based MGs with ideal thermodynamic properties  $T_x$  and  $T_g$ . All of the machine-learning models mentioned in the above works are trained based on experimental datasets. Because of the inherent disorder in atomic structures of MGs, no high-throughput computational data for amorphous materials are available, making it difficult to utilize simulation data for model training. Consequently, issues related to data distribution and sample imbalance within the experimental datasets become common challenges affecting the quality of model predictions.

The data imbalance manifests in two aspects: one is the imbalanced distribution of data related to elements; the other is the imbalance in the distribution range of performance data since extremely exceptional performance is generally found in a minority of systems. In order to provide a more specific illustration of the data imbalance, we conduct an analysis of the data distribution within a database of MGs by combining datasets from published articles [30–32] and experimental data in our research group [13,33,34]. Taking the values of  $D_{\max}$  for MGs as an example, first, as shown in Fig. 1(a), most of the common alloying elements on the periodic table are covered, among which there are more data of Mg, Al, Fe, Co, Ni, Cu, and Zr elements. Second, as shown in Fig. 1(b), more than half of the overall database consists of ribbonlike samples, while crystal data account for less than one-quarter and bulk amorphous data are the least, less than one-sixth. There is a shortage of data on BMGs due to the unclear mechanisms of GFA as well as the lack of efficient techniques for the design and preparation of BMGs. Besides, many experiments that fail to form amorphous structures are not reported, resulting in a relatively smaller amount of crystal data included. In Fig. 1(c), the statistical graph of  $D_{\max}$  for BMGs data in the database indicates that data with  $D_{\max}$  larger than 5 mm

are generally less abundant compared to data with  $D_{\max}$  less than 5 mm.

The data imbalance also occurs in each specific system, such as a system with only sparse experiment data points in the entire chemical space but dense amorphous data points near a certain composition. Models trained on imbalanced data in such systems tend to bias prediction results towards the data-rich amorphous composition region. However, in areas with no data, the lack of information does not necessarily indicate the inability to form BMGs. Moreover, it is well known that systems with similar elemental composition often show proximate high-GFA regions, which means that the data imbalance within the similar systems affects the prediction of target systems as well. Therefore, the currently available experimental datasets unavoidably suffer from imbalance issues from the elemental distribution and label-value distribution within the entire dataset, as well as the nonuniform data distribution in each specific system. When making predictions based on the imbalanced dataset, the results tend to be trapped in some area and the generalization ability is reduced, causing a missing opportunity to discover new composition with high GFA.

Optimization attempts have been made in several research studies to address the issue of imbalanced data of MGs. On the one hand, data enhancement is a good solution for imbalanced label distribution [35]. By adjusting the data sampling method, a uniform distribution of label values can be achieved. On the other hand, some scholars introduce a large number of high-throughput experimental data to dilute the imbalanced feature distribution caused by traditional experimental data [36,37]. However, current high-throughput synthesis of metallic glasses mainly focuses on film samples by magnetron sputtering, which cannot provide the quantitative measure of GFA needed in the regression problem. The impact of training-dataset selection on the model regression prediction ability in BMGs is important but there is still no systematic and comprehensive study of it.

In this work, we focus on the effect of the data imbalance on the model prediction and attempt to provide solutions to avoid the results from trapping into the intensive training-data area or from inducing by the data distribution of systems with similar composition. Ternary systems are chosen for study because they offer abundant data and avoid significant increasing complexity in chemical space compared with higher-order multicomponent systems. A machine-learning regression algorithm that integrates multiple models to achieve predictions of the  $D_{\max}$  of MGs is established. And, a series of training subdatasets containing different quantities of similar composition data for prediction are designed and their results are tested. By evaluating the prediction results of different subdatasets, the strategy to prune the training dataset to model and predict various systems commencing from a similar information base is proved to mitigate the impact of data imbalance. To compensate the information loss caused by data reduction, we introduce an experimental iterative approach. Using this workflow, we validate its effectiveness through identifying the high-GFA composition in the Zr-Al-Cu system by treating it as a completely unknown system for prediction. Finally, the application of this workflow in the field of BMGs and the existing challenges are discussed.

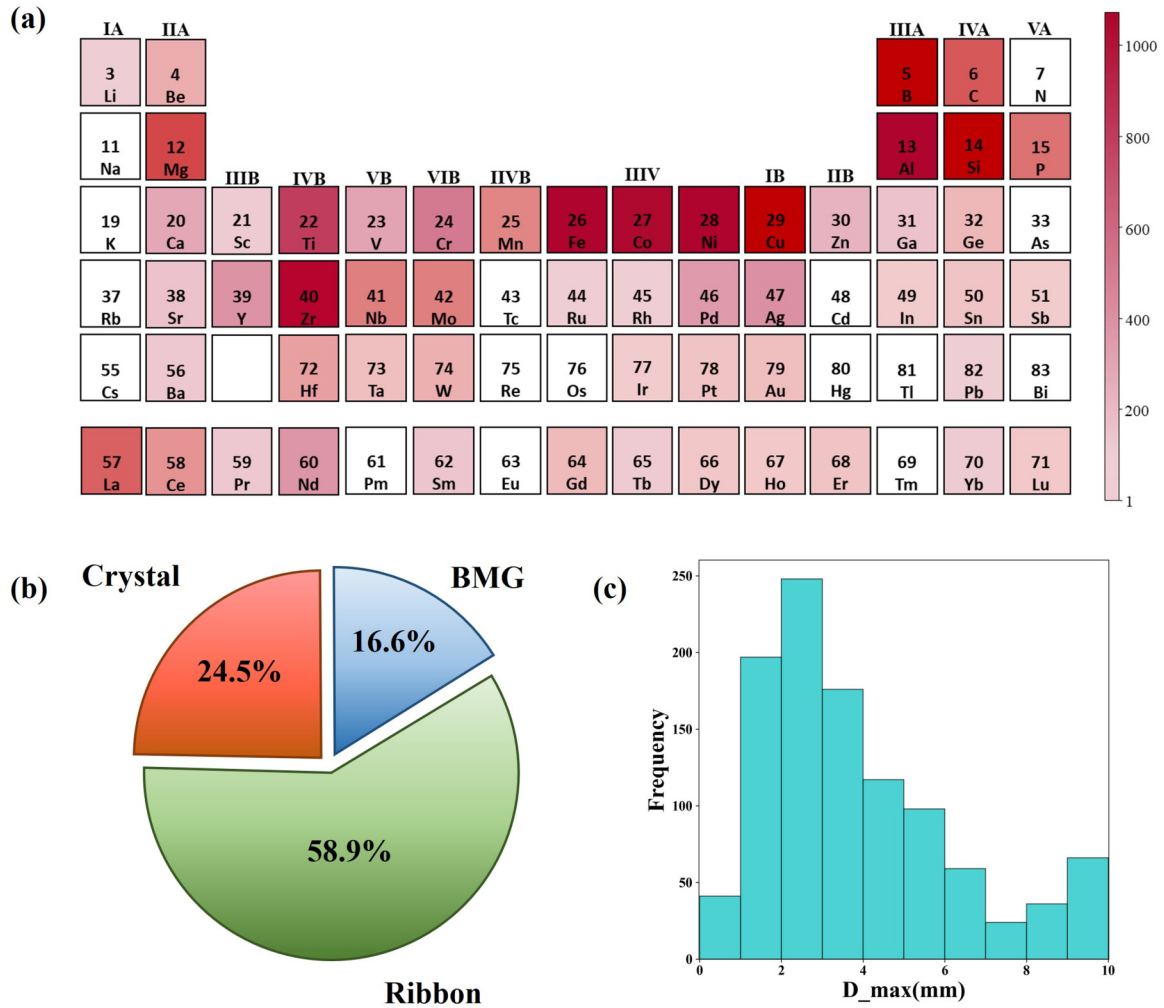


FIG. 1. Distribution analysis of GFA datasets (a) Statistics of element occurrence frequencies in the datasets. (b) Pie chart of label value types. The “Crystal” label represents data that can only be synthesized as crystals, “Ribbon” represents data corresponding to ribbon samples obtained by melt spin quenching, and “BMG” represents data that can be synthesized as BMGs. (c) Statistical chart of  $D_{\max}$  of BMGs in the datasets.

## II. METHOD

### A. General workflow and model training

As shown in Fig. 2, we export five groups of descriptors using the MATMINER PYTHON library [38] to represent the physical and chemical properties of each composition. On this dataset, we initially train a set of models, including Support Vector Machine(SVM), Random Forest (RF), eXtreme Gradient Boosting (XGBoost), and Light Gradient Boosting (LightGBM). These models are classical machine-learning models, which have good prediction ability for MGs and

have certain anti-overfitting ability [39–41]. Based on the HYPEROPT library, we choose the Tree-Structured Parzen Estimator optimization algorithm and stop iterate when there is no improvement on the test set for 50 consecutive iterations in the fivefold validation [42]. The selection range of hyperparameters is detailed in Table S1; see the Supplemental Material [43]. Afterwards, we perform model fusion operation by using the four predicted values of submodels as inputs and optimizing the weights of these four values as hyperparameters. The final prediction result can be written as

$$\text{pred} = \frac{x_1 \times y_{\text{SVM}} + x_2 \times y_{\text{RF}} + x_3 \times y_{\text{XGBoost}} + x_4 \times y_{\text{LightGBM}}}{x_1 + x_2 + x_3 + x_4}, \quad (1)$$

where  $x$  is the weight of each submodel prediction result, which is optimized as a hyperparameter.  $y$  represents the prediction result output by each submodel, and  $\text{pred}$  represents the result of the final output of the fusion model.

When the whole dataset is chosen for training, compared to the four individual models, the fusion model shows better prediction performance, with an  $R^2$  value of 0.82 on the test set under the condition of fivefold cross validation (as shown

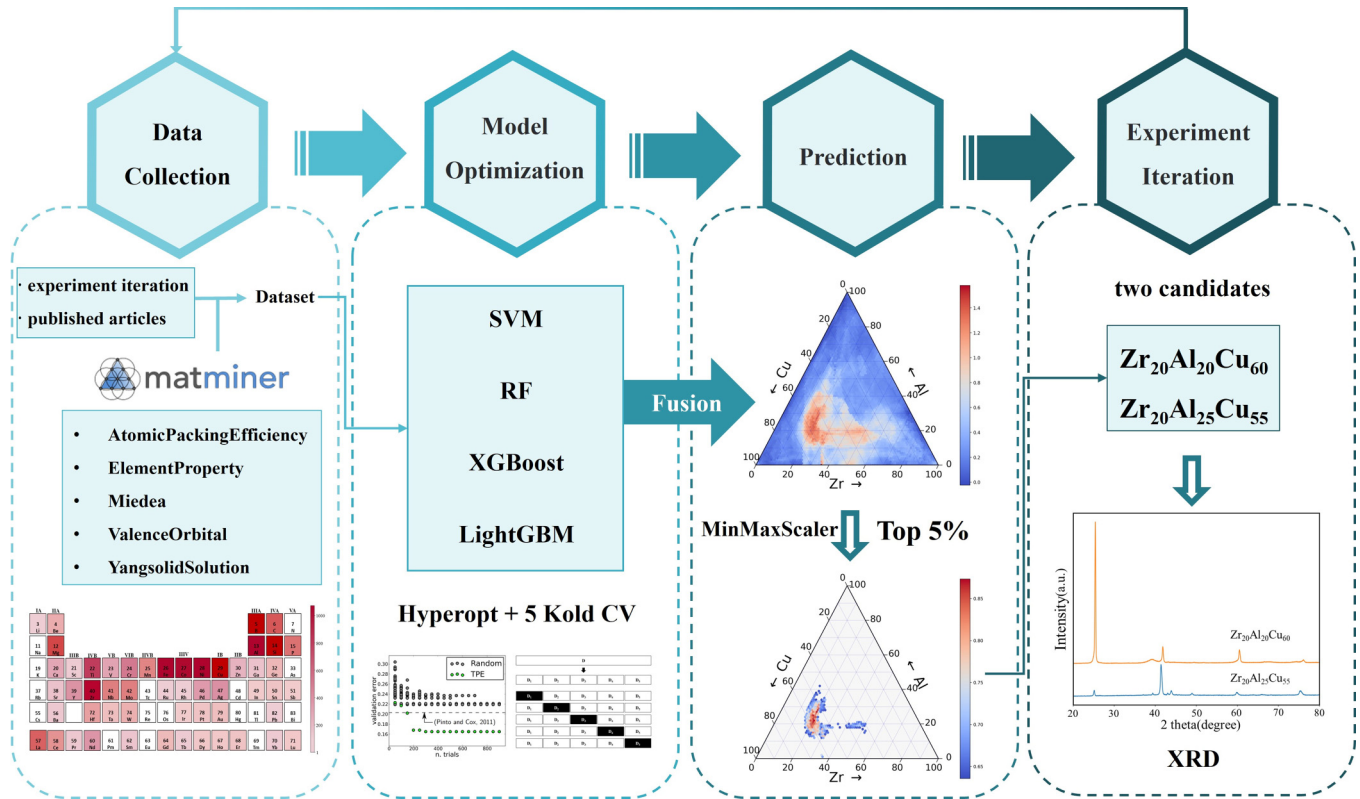


FIG. 2. Machine learning and experimental iteration workflow proposed in this work. Models are trained and optimized using HYPEROPT, with model fusion improving prediction performance. The top 5% points exhibiting high GFA are selected based on ranked  $D_{\max}$  prediction. Two candidates are experimentally validated in each iteration process and their results are added to the training set for the next iteration.

in Fig. S1; see the Supplemental Material [43]). For each ternary system, GFA predictions are conducted by varying the composition with a 1% atomic ratio change, resulting in a total of 4851 composition predicted for each system. We normalize the predicted  $D_{\max}$  values to the range of [0,1], where a value closer to 1 indicates a higher  $D_{\max}$  prediction value in this system, while a value closer to 0 indicates a lower  $D_{\max}$  prediction value. The points corresponding to the top 5% highest predicted  $D_{\max}$  values of each system are selected as candidate composition and illustrated in the phase diagram. In the largest candidate area, we choose two candidates to synthesize in each experimental iteration for validation: one is the point with the highest predicted value and the other has the largest predicted value in a group of six points surrounding the first one. We prepare ribbon samples by melt spinning, and prepare cylindrical BMG samples of different sizes by copper mold casting in molds of different diameters, starting from 1 mm diameter and increasing by 0.5 mm each time until the sample exhibits diffraction patterns containing crystalline structures in XRD testing. The obtained experimental results are subsequently incorporated as training data into the dataset for the next iteration.

### B. Training-data processing

The  $D_{\max}$  of the composition that can be prepared as BMGs is taken as the original reported or experimental value; the  $D_{\max}$  of the composition that can be prepared as ribbon amorphous samples is taken as 0.2 mm [44], and the corresponding

label value of the composition that cannot be obtained as amorphous samples is taken as 0 mm. Different reported values of the same composition are averaged and included in the database. Afterwards, the entire dataset will be pruned step by step by removing part of data that contributes to different forms of data imbalance, to test how to minimize the impact of data imbalance on the model's prediction ability.

### C. Descriptor selection

For feature selection, we utilize the MATMINER PYTHON library [38] to derive descriptors based on material composition. We incorporate descriptors from "ElementProperty" database and "ValenceOrbital" database to calculate the physical and chemical properties of each element, such as electronegativity, atomic radius, thermal conductivity, electron distribution, and so on. Additionally, we also consider other characteristics that may be related to GFA. Laws *et al.* [45] investigated the correlation between mean packing efficiency and GFA in their study, revealing that ideal atomic packing efficiency is likely to be one of the necessary conditions for high GFA. Therefore, based on their studies, we derive descriptors in this group to describe the atomic packing efficiency according to alloy compositions. We also derive solid-solution phase-stability descriptors based on Yang's work [46], in which multicomponent BMGs are concentrated in the range of  $\Omega$  (the entropy of mixing timing the average melting temperature of the elements over the enthalpy of mixing) less than 1.1 and  $\gamma$  (the mean-square deviation



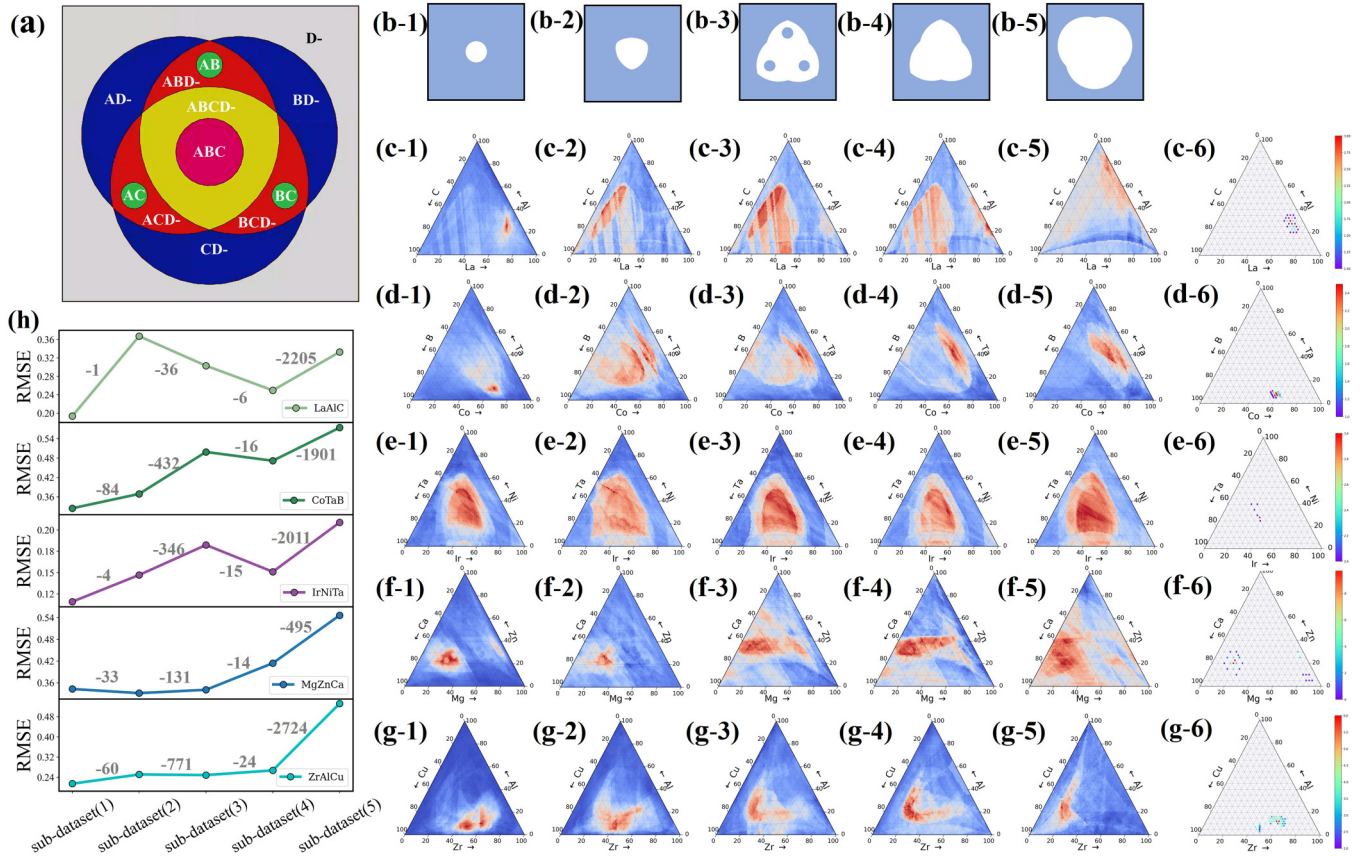


FIG. 3. The impact of varying degrees of data imbalance on the model prediction: (a) diagram depicting the data-type division in a ternary system; (b) diagrams illustrating the data range of the subdatasets; (c)–(g) prediction results of La-Al-C, Co-Ta-B, Ir-Ni-Ta, Mg-Zn-Ca, Zr-Al-Cu, system [( $x - 1$ ) ~ ( $x - 5$ )] represent the prediction results in the cases of (1) to (5) subdatasets in turn and ( $x - 6$ ) represents the true value of the experimental reports, where  $x = b, c, d, e,$  and  $f$ ]; (h) effect of pruning relevant data on the error of model prediction results (the labeled numbers indicate the data entries reduced from the previous dataset).

of the atomic size of elements) less than 5%, so we derive these two descriptors based on this paper. Finally, we derive three descriptors based on the energy difference between crystalline and amorphous states using the semiempirical model of Miedema *et al.* [47]. A total of 86 descriptors are exported. Among them, 39 descriptors from the Element-Property database are excluded due to their lack of variation with changes in elemental composition ratios within the same system. In addition, 6 descriptors with missing values are removed, and 12 descriptors are further removed due to strong linear correlations. Consequently, a refined dataset consisting of 29 descriptors (as shown in Table S2; see the Supplemental Material [43]) is obtained through these filtration steps. The absolute value of the linear correlation between all pairwise descriptors is less than 0.8 (as shown in Fig. S2; see the Supplemental Material [43]), which means that there is no significant interdependence between this set of descriptors and they can be used for the prediction of  $D_{\max}$ .

### III. RESULTS AND DISCUSSION

#### A. Effect of imbalanced training datasets on prediction

Although, as most of the reported machine-learning models of BMGs, our optimized fusion model trained on the entire experimental dataset demonstrates prediction ability on

the test set under the condition of fivefold cross validation, its ability in systems which have not been experimentally investigated or in the composition areas which have sparse data is still questionable. Therefore, based on the aforementioned dataset, we design a series of subdatasets to emulate the varying degrees of data loss and to explore the influence of data imbalance on the predictive ability of the model. Subdatasets for the prediction target of  $A-B-C$  system (where  $A, B,$  and  $C$  represent different elements) are designed and their prediction results are evaluated in Fig. 3. As shown in Fig. 3(a), we classify the data containing varying degrees of information on the composition of  $A-B-C$  system into five categories as follows: (i) data for composition containing only one of the three elements, denoted as  $AD-$ ,  $BD-$ , and  $CD-$  (“ $D$ ” represents elements other than  $A, B,$  and  $C$ , and “ $D-$ ” can represent more than one element); (ii) binary data containing two of the three elements, namely  $AB$ ,  $BC$ , and  $AC$ ; (iii) data for alloys of three or more elements containing two of the three elements, namely  $ABD-$ ,  $BCD-$ , and  $ACD-$ ; (iv) data for alloys of four or more elements containing these three elements, i.e.,  $ABCD-$ ; and (v) data representing the ternary alloy system itself, namely  $ABC$ . For example, in the case of the Zr-Al-Cu system as the target,  $Ti_{45}Zr_{20}Be_{35}$ ,  $Zr_{50}Cu_{50}$ ,  $Zr_{40}Al_{20}Zn_{40}$ ,  $Zr_{40}Al_{25}Cu_{25}Co_{10}$ , and  $Zr_{60}Al_{10}Cu_{30}$  belong to the data category (i), (ii), (iii), (iv), and (v), respectively.

From categories (i) to (v), the data exhibit an increasing association with the target system. To explore their effect on the model prediction, five subdatasets with varying degrees of missing relevant data are designed by reducing the data in category from (v) to (i) step by step, as shown in Fig. 3(b). In this way, the imbalance decreases from left to right for the datasets presented in Fig. 3(b). Five ternary systems, La-Al-C [34], Co-Ta-B [33,48,49], Ir-Ni-Ta [50], Mg-Zn-Ca [51], and Zr-Al-Cu [52–55] are trained based on the above five-category subdatasets. The model's prediction ability is compared among them and the results are shown in Figs. 3(c)–3(g). To accurately assess the impact of data imbalance, we normalize the collected experimental data in each prediction to [0, 1]. The prediction ability is quantified through the relative mean-squared error (RMSE), as presented in Fig. 3(h).

It is evident that the model's predicted results in subdataset (1) show a significant overlap with the high-GFA region reported in collected experimental data. In Fig. 3(h), predicted results in subdatasets (1) generally have the smallest RMSE. This is mainly because the *ABCD*-type data present in subdatasets (1) are generally developed from the *ABC* amorphous composition; thus, they have a strong similarity to the target system. In the La-Al-C system, a significant increase of the RMSE is observed with the deletion of only one data point, as shown in Figs. 3(c-1) and 3(c-2). The composition corresponding to this data point is  $\text{La}_{60}\text{Al}_{26}\text{Ga}_4\text{C}_{10}$ , where the Ga element is adjacent to Al in the periodic table and this data point is therefore close to  $\text{La}_{60}\text{Al}_{30}\text{C}_{10}$  in feature space, playing a strong guiding impact to trap the predicted high- $D_{\max}$  value near to this composition. In the subsequent data gradient change, the regions of high-GFA composition in the predicted results of La-Al-C and Co-Ta-B systems are almost unchanged, as shown in Figs. 3(c) and 3(d). For Ir-Ni-Ta system, as shown in Fig. 3(e), the variation of high-GFA regions is insignificant in several cases. Nevertheless, for the Mg-Zn-Ca and Zr-Al-Cu systems the impact of data imbalance become notable starting from subdataset (4), where the RMSE increases obviously as shown in Fig. 3(h). The inclusion of category (iii) data in the training set causes the predicted high-GFA regions aggregating to the experiment reported high-GFA regions, resulting in the prediction results being biased towards the data-intensive regions in the training dataset. The bias caused by the data imbalance poses challenges for the discovery of BMG compositions. For example, in the Mg-Zn-Ca system, there are two high-GFA compositions with significant differences in composition. To be specific, both  $\text{Mg}_{20}\text{Zn}_{20}\text{Ca}_{60}$  in the low-Mg region and  $\text{Mg}_{70}\text{Ca}_5\text{Zn}_{25}$  in the high-Mg region have the same  $D_{\max}$  value. However, in Fig. 3(g-2), only one high- $D_{\max}$  region can be observed, and it is only after removing category (iii) data that a second high- $D_{\max}$  prediction region can be observed in Fig. 3(f-3). This imbalance in the data is likely to result in the overlooking of potential high-GFA regions, thus significantly compromising the effectiveness of machine learning on looking for new composition with candidate properties.

To mitigate the undesirable influence of a large amount of relevant experimental data on the prediction results, we choose the training-set selection strategy of subdataset (3) for further discussion. In this final selected subdataset, there are

no data that are highly similar to the element composition of the ternary system to be predicted. Taking the Zr-Al-Cu system as an example, the subdataset used at this time does not contain the ternary data containing two of the three elements of Zr, Al, and Cu, which reflects the unbiasedness. This selection criterion serves a dual purpose. On one hand, it helps reduce the adverse guiding effect arising from similar composition data. On the other hand, it enables us to model composition exploration in unfamiliar systems, where data category (iii) to (iv) for the target system are unavailable. Thus, each system is predicted as an unexplored system. To compensate for the information loss resulting from data reduction, we next conduct a small number of experiments based on the prediction results and iterate the experimental data into the machine-learning model for another round of predictions, aiming to identify the high-GFA area in the target system even if it is regarded as an entirely unexplored system [28,56].

## B. Experimental iteration

We choose Zr-Al-Cu system to validate the iteration scheme between experiments and machine-learning models, due to its relatively simple preparation process, low cost, as well as abundant experimental data available for comparison. As shown in Fig. 4(a), the circled pentagrams correspond to the data points that are experimentally validated and used for each iteration. Figure 4(b-1) shows the top-5% region of the  $D_{\max}$  in the initial model prediction without any iteration, where the color represents the relative predicted value. Figures 4(b-2)–4(b-4) display the prediction results after the first, second, and third iterations, respectively. According to the prediction results in Fig. 4(b-1), two candidates,  $\text{Zr}_{20}\text{Al}_{20}\text{Cu}_{60}$  and  $\text{Zr}_{20}\text{Al}_{25}\text{Cu}_{55}$  with a difference of 5 at. % in composition are chosen for experimental verification and iteration. We explore the  $D_{\max}$  of candidates by copper casting and melt-spin quenching. The experimental results (see Fig. S3, the Supplemental Material [43]) indicate that  $\text{Zr}_{20}\text{Al}_{20}\text{Cu}_{60}$  and  $\text{Zr}_{20}\text{Al}_{25}\text{Cu}_{55}$  cannot be prepared as amorphous samples, and the XRD patterns of the ribbon samples obtained through melt-spin quenching exhibit obvious crystalline structures. The two experimental data are added into the training set and another model is obtained. Using this model, the high-GFA area moves to regions with higher Zr concentration as shown in Fig. 4(b-2). In the second iteration,  $\text{Zr}_{40}\text{Al}_{20}\text{Cu}_{40}$  and  $\text{Zr}_{45}\text{Al}_{20}\text{Cu}_{35}$  are chosen for experimental validation, and ribbon amorphous samples are successfully yielded for them. Furthermore, Fig. 4(b-3) is the result predicted by the next trained model with two more experimental data added. The third iteration data,  $\text{Zr}_{50}\text{Al}_5\text{Cu}_{45}$  and  $\text{Zr}_{55}\text{Al}_5\text{Cu}_{40}$ , can obtain cylindrical amorphous samples with diameters of 2.5 and 1.5 mm, respectively. As depicted in Fig. 4(b-4), the highest point of the predicted result coincides with the experimental iteration point, and is close to the experimental reported highest  $D_{\max}$  at composition within the distance of 10 at.%. It can be observed that as we add experimental data, the prediction results gradually approach the amorphous area. During the experimental process, the GFA corresponding to the selected iteration data becomes higher and higher, confirming that iteration can indeed improve the prediction effectiveness of the model.

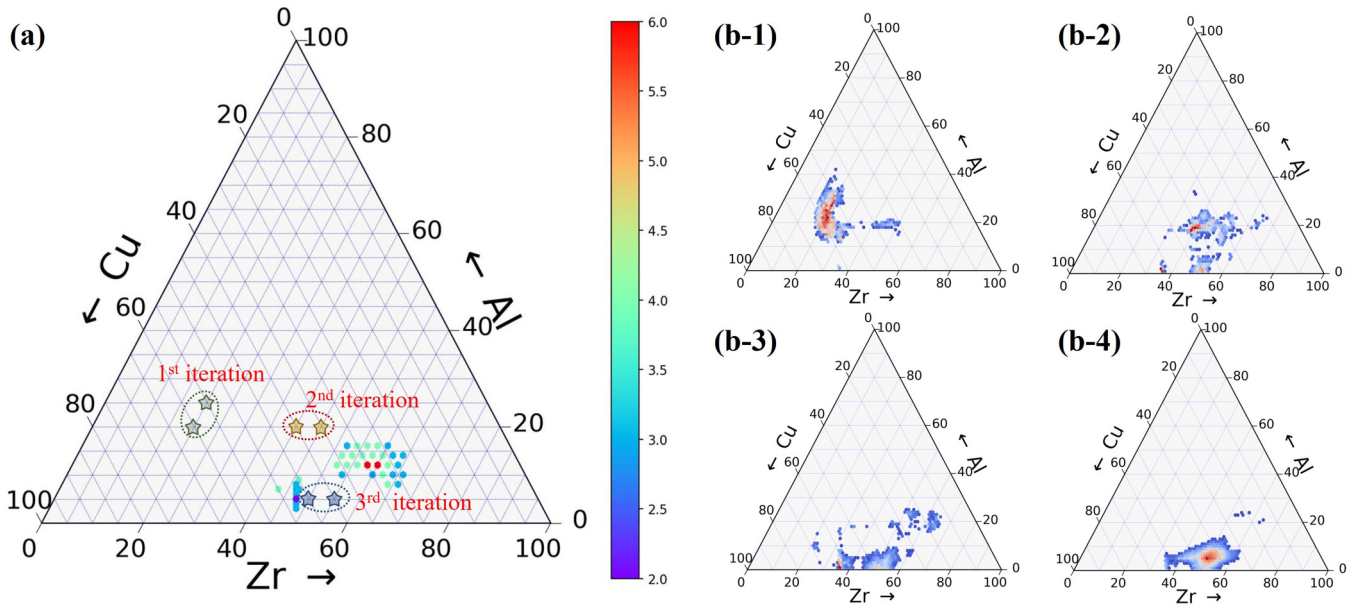


FIG. 4. Iterative prediction empowered by adding experimental data step by step. (a) reported  $D_{\max}$  for Zr-Al-Cu system, where pentagrams correspond to the iteration data points; (b) top 5% iterative prediction results for the Zr-Al-Cu system, showing initial results without iteration (b-1) and after the first (b-2), second (b-3), and third (b-4) iterations.

Additionally, we examine the high-GFA region centered around  $Zr_{50}Al_5Cu_{45}$  in Fig. 4(b-4). We proceed to select boundary points with equal Zr content, equal Al content, and equal Cu content, respectively, using increments of five atomic ratios to conduct experiments in order to verify the potential formation of BMGs in this area. As shown in Fig. S3 (see Fig. S3, the Supplemental Material [43]), the 1 mm diameter rodlike samples of  $Zr_{45}Al_{10}Cu_{45}$  and  $Zr_{35}Al_5Cu_{60}$  prove to be amorphous; the other two candidates can be prepared into ribbon amorphous samples. This indicates that the pruned dataset effectively mitigates the data imbalance issue and iteration scheme with a small amount of experiment exploration that help to identify a relatively high-GFA region in Zr-Al-Cu systems, even if there are no data of the ternary system provided beforehand. Training and predicting each system as an unexplored one by reducing the strong related data from the dataset decrease the chance to be trapped into the known data-intensive area and will benefit to find more composition with high GFA.

#### IV. CONCLUSION

In this work, we analyze the origin and effect of data imbalance in the datasets of BMGs. The way to mitigate the effect of this imbalance is proposed by testing the model prediction ability using varying pruned degrees of the datasets. And, the subset without *ABC*-related information is chosen since it avoids the results trapped at the data-intensive area. By combining the machine learning and experimental iteration, we mitigate the negative impact of data imbalance and accelerate the search for high-GFA regions, especially suitable for discovery of composition in unexplored system and beyond experimental intensive data area in known systems.

The main conclusions are as follows:

(1) By designing multiple subdatasets with gradients of similarity to the elements of the target system, we demonstrate that imbalanced data distribution can introduce bias, which has a strong guiding impact on the predictive results of the model. An imbalanced dataset can result in biased predictions in densely populated areas, reducing the applicability of the model. By restricting the training sets to exclude the *ABC*-related experimental data, all the target system can be predicted at almost the same level, no matter how many parts of the specific chemical space have not been studied yet. It is important to note that this does not negate the effectiveness of adding data in an even manner, as doing so can improve the prediction accuracy of the model to some extent. It is suitable for integration with high-throughput BMGs preparation methods such as laser melting [57–59] to speed up the high-quality data production in BMGs exploration.

(2) We propose a workflow that involves a small number of experiments and iterative model training to identify high-GFA regions in a specified system. This combination of machine learning and experiment methods can help researchers to identify regions with relatively higher GFA in the Zr-Al-Cu system. Additionally, this method can be expanded to predict other properties of amorphous alloys, effectively mitigating the impact of data imbalance and reducing the time required for experimentation. By simply substituting the labels in the dataset, we can extend the predictive target to other properties of interest in the field of MGs, such as glass transition temperature or elastic modulus. The success of the above strategy lies in two aspects. First, by excluding the *ABC*-related data from the training dataset, the prediction values are prevented from being biased towards the dense region of training data, which mostly only contains amorphous cases. Second, the data iteratively added through experiments include both crystalline and amorphous results,



continuously improving the prediction across the whole chemical space.

Theoretically, the computational and experimental costs increase when targeting a quaternary or higher system. The composition space to be explored becomes more intricate, leading to a broader range of iteration choices. Nevertheless, when compared with the conventional trial and error approach for exploring compositions, the method outlined in this paper demonstrates notable efficiency advantages within the constraints of comparable complexity. This approach will expedite the development of BMG compositions, especially for unexplored systems.

All the data and codes for the fusion model can be accessed via GitHub [60].

## ACKNOWLEDGMENTS

This study was funded by the National Natural Science Foundation of China (Grants No. 52171150, No. 52172258, No. 52022106), and the National Key Research and Development Program of China (Grant No. 2018YFA0703601).

R.L. designed this work; R.L. and R.X. guided the completion of the method; X.G. and R.X. constructed the models; X.G. conducted the materials synthesis and model iterations; and J.B. and X.L. collected and organized the dataset. All the authors participated in the analysis of the data and discussions of the results, as well as in preparing the paper.

All authors declare no financial or nonfinancial competing interests.

- 
- [1] A. Rajkomar, J. Dean, and I. Kohane, *N. Engl. J. Med.* **380**, 1347 (2019).
- [2] D. Gunduz, P. de Kerret, N. D. Sidiropoulos, D. Gesbert, C. R. Murthy, and M. van der Schaar, *IEEE J. Sel. Areas Commun.* **37**, 2184 (2019).
- [3] Y. Wang, S. Wu, W. Shao, X. Sun, Q. Wang, R. Xiao, and H. Li, *J. Materiomics* **8**, 1038 (2022).
- [4] J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, *npj Comput. Mater.* **5**, 83 (2019).
- [5] G. L. W. Hart, T. Mueller, C. Toher, and S. Curtarolo, *Nat. Rev. Mater.* **6**, 730 (2021).
- [6] L. Burzawa, S. Liu, and E. W. Carlson, *Phys. Rev. Mater.* **3**, 033805 (2019).
- [7] A. J. Gormley and M. A. Webb, *Nat. Rev. Mater.* **6**, 642 (2021).
- [8] J. F. Masson, J. S. Biggins, and E. Ringe, *Nat. Nanotechnol.* **18**, 111 (2023).
- [9] H. Xin, *Nat. Energy* **7**, 790 (2022).
- [10] Z. Fan, E. Ma, and M. L. Falk, *Phys. Rev. Mater.* **6**, 065602 (2022).
- [11] Y. Liu, T. Zhao, W. Ju, and S. Shi, *J. Materiomics* **3**, 159 (2017).
- [12] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, *Nature (London)* **559**, 547 (2018).
- [13] J. Bi, X. Wei, X. Liu, R. Li, R. Xiao, and T. Zhang, *Scr. Mater.* **228**, 115336 (2023).
- [14] J. Zhou, Q. Wang, X. Hui, Q. Zeng, Y. Xiong, K. Yin, B. Sun, L. Sun, M. Stoica, W. Wang, and B. Shen, *Mater. Des.* **191**, 108597 (2020).
- [15] X. Zhang, Y. Yang, Y. Liu, Z. Jia, Q. Wang, L. Sun, L. C. Zhang, J. J. Kruzic, J. Lu, and B. Shen, *Adv. Mater.* **35**, e2303439 (2023).
- [16] C. Ma, N. Nishiyama, and A. Inoue, *Mater. Trans.* **43**, 1161 (2002).
- [17] E. S. Park and D. H. Kim, *J. Mater. Res.* **19**, 685 (2003).
- [18] S. F. Zhao, N. Chen, P. Gong, and K. F. Yao, *J. Alloys Compd.* **647**, 533 (2015).
- [19] Y. Li, J. Ma, P. K. Liaw, and Y. Zhang, *J. Alloys Compd.* **913**, 165294 (2022).
- [20] J. Lai, W. Hu, A. Datye, J. Liu, J. Schroers, U. D. Schwarz, and J. Yu, *J. Mater. Sci. Technol.* **84**, 97 (2021).
- [21] S. A. Kube, S. Sohn, R. Ojeda-Mota *et al.*, *Nat. Commun.* **13**, 3708 (2022).
- [22] H. W. Yoon, H. C. Lee, S. Y. Shin, S. H. Kwon, and K. I. Moon, *Coatings* **10**, 398 (2020).
- [23] Z. P. Lu, H. Tan, Y. Li, and S. C. Ng, *Scr. Mater.* **42**, 667 (2000).
- [24] Z. P. Lu and C. T. Liu, *Acta Mater.* **50**, 3501 (2002).
- [25] X. H. Du, J. C. Huang, C. T. Liu, and Z. P. Lu, *J. Appl. Phys.* **101**, 086108 (2007).
- [26] X. W. Liu, Z. L. Long, W. Zhang, and L. M. Yang, *J. Alloys Compd.* **901**, 163606 (2022).
- [27] D. You, H. Zhang, S. Ganorkar, T. Kim, J. Schroers, J. J. Vlassak, and D. Lee, *Acta Mater.* **231**, 117861 (2022).
- [28] F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers, and A. Mehta, *Sci. Adv.* **4**, eaaq1566 (2018).
- [29] J. Jeon, G. Kim, N. Seo, H. Choi, H.-J. Kim, M.-H. Lee, H.-K. Lim, S. B. Son, and S.-J. Lee, *J. Mater. Res. Technol.* **16**, 129 (2022).
- [30] L. Ward, A. Agrawal, A. Choudhary, and C. Wolverton, *npj Comput. Mater.* **2**, 16028 (2016).
- [31] Z. Long, H. Wei, Y. Ding, P. Zhang, G. Xie, and A. Inoue, *J. Alloys Compd.* **475**, 207 (2009).
- [32] Z. P. Lu, H. Bei, and C. T. Liu, *Intermetallics* **15**, 618 (2007).
- [33] J. Wang, R. Li, N. Hua, and T. Zhang, *J. Mater. Res.* **26**, 2072 (2011).
- [34] L. Zhang, R. Li, T. Xu, H. Zhang, and T. Zhang, *Intermetallics* **52**, 92 (2014).
- [35] X. Liu, Z. Long, W. Zhang, L. Yang, and Z. Li, *J. Noncryst. Solids* **621**, 122645 (2023).
- [36] Z. Liu, C. Chen, Y. Zhou, L. Zhang, and H. Wang, *Scr. Mater.* **229**, 115347 (2023).
- [37] Y. Yao, T. Sullivan, F. Yan, J. Gong, and L. Li, *Scr. Mater.* **209**, 114366 (2022).
- [38] L. Ward, A. Dunn, A. Faghaninia, N. E. R. Zimmermann, S. Bajaj, Q. Wang, J. Montoya, J. Chen, K. Bystrom, M. Dylla, K. Chard, M. Asta, K. A. Persson, G. J. Snyder, I. Foster, and A. Jain, *Comput. Mater. Sci.* **152**, 60 (2018).
- [39] X. Li, G. Shan, and C. H. Shek, *J. Mater. Sci. Technol.* **103**, 113 (2022).
- [40] X. Liu, Z. Long, and L. Peng, *J. Noncryst. Solids* **602**, 122095 (2023).
- [41] L. Zhang, H. Chen, X. Tao, H. Cai, J. Liu, Y. Ouyang, Q. Peng, and Y. Du, *Mater. Des.* **193**, 108835 (2020).



- [42] J. Bergstra, D. Yamins, and D. Cox, in *Proceedings of the 30th International Conference on Machine Learning*, edited by D. Sanjoy and M. David (PMLR, Proceedings of Machine Learning Research, 2013), p. 115.
- [43] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevMaterials.8.055602> for details of machine-learning process (Table S1, Table S2, Fig. S1, and Fig. S2) and XRD patterns of the experimental iterative data (Fig. S3).
- [44] L. Ward, S. C. O’Keeffe, J. Stevick, G. R. Jelbert, M. Aykol, and C. Wolverton, *Acta Mater.* **159**, 102 (2018).
- [45] K. J. Laws, D. B. Miracle, and M. Ferry, *Nat. Commun.* **6**, 8123 (2015).
- [46] X. Yang and Y. Zhang, *Mater. Chem. Phys.* **132**, 233 (2012).
- [47] A. R. Miedema, P. F. de Châtel, and F. R. de Boer, *Physica B+C* **100**, 1 (1980).
- [48] J. Wang, I. Kaban, V. Levytskyi, R. Li, J. Han, M. Stoica, R. Gumenuk, and K. Nielsch, *J. Alloys Compd.* **860**, 158398 (2021).
- [49] X.-Q. Wei, J.-Z. Bi, and R. Li, *Acta Phys. Sin.* **66**, 176408 (2017).
- [50] M. X. Li, S. F. Zhao, Z. Lu, A. Hirata, P. Wen, H. Y. Bai, M. Chen, J. Schroers, Y. Liu, and W. H. Wang, *Nature (London)* **569**, 99 (2019).
- [51] O. N. Senkov and J. M. Scott, *J. Noncryst. Solids* **351**, 3087 (2005).
- [52] Q. He, Y.-Q. Cheng, E. Ma, and J. Xu, *Acta Mater.* **59**, 202 (2011).
- [53] P. Jia, Z.-d. Zhu, E. Ma, and J. Xu, *Scr. Mater.* **61**, 137 (2009).
- [54] T. L. Cheung and C. H. Shek, *J. Alloys Compd.* **434-435**, 71 (2007).
- [55] J. Eckert, J. Das, K. B. Kim, F. Baier, M. B. Tang, W. H. Wang, and Z. F. Zhang, *Intermetallics* **14**, 876 (2006).
- [56] Z. Y. Rao, P. Y. Tung, R. W. Xie, Y. Wei, H. B. Zhang, A. Ferrari, T. P. C. Klaver, F. Körmann, P. T. Sukumar, A. K. da Silva, Y. Chen, Z. M. Li, D. Ponge, J. Neugebauer, O. Gutfleisch, S. Bauer, and D. Raabe, *Science* **378**, 78 (2022).
- [57] Y. Lu, Y. Huang, J. Wu, X. Lu, Z. Qin, D. Daisenberger, and Y.-L. Chiu, *Intermetallics* **103**, 67 (2018).
- [58] Y. Luo, L. Xing, Y. Jiang, R. Li, C. Lu, R. Zeng, J. Luo, P. Zhang, and W. Liu, *Materials (Basel)* **13**, 597 (2020).
- [59] X. Liu, J. Bi, Z. Meng, R. Li, Y. Li, and T. Zhang, *Tribol. Int.* **162**, 107142 (2021).
- [60] [https://github.com/xuhegg/ML\\_BMG\\_GFA](https://github.com/xuhegg/ML_BMG_GFA).