npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-025-01753-9

Artificial intelligence can recognize metallic glasses in vast compositional space with sparse data

Weijie Xie[1,2,4], Yitao Sun[1,4], Chao Wang[1], Mingxing Li[1], Fucheng Li[1] & Yanhui Liu[1,3]

Glass formation is frequently observed in metallic alloys. Machine learning has been applied to discover new metallic glasses. However, the incomplete understanding of glass formation hinders descriptor selection and material property representation. Here, we use X-ray diffraction spectra, the essential tool for identifying amorphous structure, as an intermediate link. By representing spectra as images, we train generative models to produce high-fidelity spectra for all alloys in multicomponent alloy systems. Training with spectra from a tiny fraction of the total alloys is sufficient for accurate spectra generation, enabling the identification of compositional regions with a high probability of glass formation. The shift from numerical to image-based representation unlocks the potential of machine learning in the design of glass-forming alloys. Furthermore, our approach is applicable to a wide range of materials and spectroscopic techniques. We anticipate that this strategy will accelerate materials discovery across previously unexplored compositional and processing spaces.

Metallic glasses (MG) are alloys characterized by the absence of long-range topological order in microscopic structure[1]. Their unique structural and functional properties, such as high strength[2], excellent wear resistance[3], and plastic-like processability[4,5], make them attractive candidates for technological usage[6]. The ability of an alloy to form a glass is usually quantified by the lowest cooling rate required for vitrification. Since the discovery of MGs[7] in 1960, extensive efforts have been made to develop alloys capable of forming glass at the lowest possible cooling rate[8]. Because only a few alloys exhibit both strong glass-forming ability and desirable properties, it has been suggested that a much broader compositional space should be explored[9]. To circumvent the complicated measurement of cooling rate, numerous indirect criteria have been proposed to predict glass-forming alloys[10,11]. However, the complexity of glass formation and the diverse selection of constituent elements in MG development make accurate prediction challenging. Even though glass formation has been suggested to be associated with atomic packing[12,13], quantitative correlation remains poorly understood[14]. As a consequence, design and development of new glass-forming alloys are primarily dependent on empirical rules followed by cumbersome experimental verifications[15]. Combinatorial methods have been deployed to accelerate the search for glass-forming alloys[15–19]. However, the compositional space defined by multiple elements in MGs is vast[9]. Current combinatorial methods can only handle a couple of elements at a time[18], making the methods insufficient to steer through the vast compositional space.

With sufficient training data and the ability to identify patterns in large datasets, machine learning (ML) approaches have been used to uncover correlations between the characteristics of MGs and their properties. These include glass formation[18,20–23], characteristic temperature[24], atomic structures[25,26], elastic moduli[27], soft magnetic properties[28,29], and prediction of alloy compositions[30]. These correlations have been used to develop predictive ML models. However, few alloys with stronger glass-forming ability than previously reported have been identified using these models[31]. This limitation arises because the descriptors in these ML models often rely on empirical criteria designed for the selection of alloy systems rather than specific alloys[32]. In addition, weight-averaged descriptor values based on elemental composition fail to capture the complex many-body interactions involved in glass formation[31,32].

In either conventional trial-and-error approaches or combinatorial methods, X-ray diffraction (XRD) is the essential tool for confirming glass formation of an alloy[15–17,19]. The diffraction spectrum of an MG is characterized by broad halo peaks, which provide valuable information about structural features and atomic packing. For example, the power-law scaling of the first diffraction peak position (q_1) with atomic volume reveals the fractal nature of atomic packing in MGs[33]. The width of the first diffraction peak (Δq) has been correlated with glass-forming ability, where a broader peak suggests more dispersed structural units and stronger glass-forming ability[17]. Thus, XRD can be considered as the bridge between microscopic

[1]Institute of Physics, Chinese Academy of Sciences, Beijing, China. [2]School of Physics, University of Chinese Academy of Sciences, Beijing, China. [3]Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing, China. [4]These authors contributed equally: Weijie Xie, Yitao Sun. e-mail: yanhui.liu@iphy.ac.cn

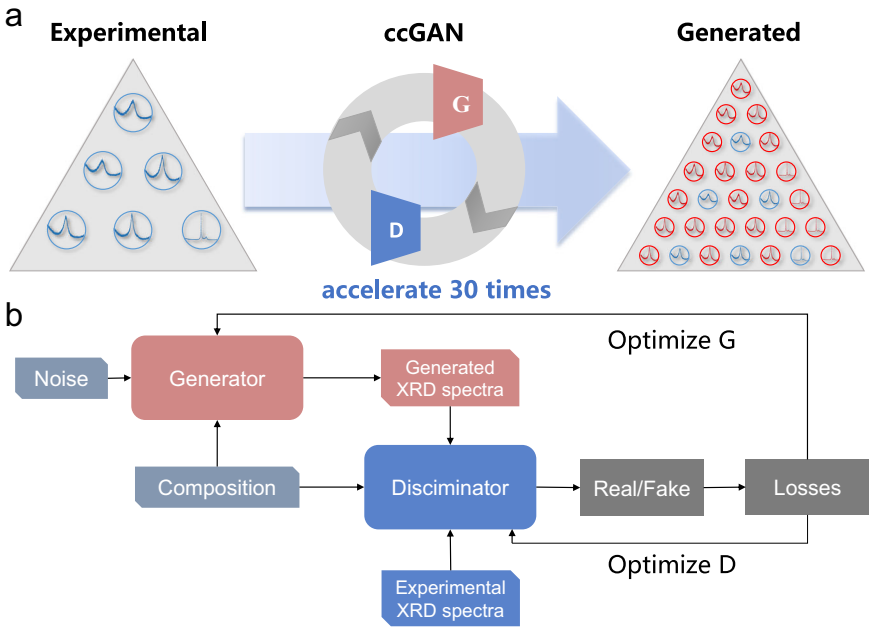

Fig. 1 | Workflow for the generation of high-fidelity XRD spectra. a The XRD spectra for alloys sparsely distributed in a multicomponent system are collected and used as training and testing sets for the model of continuous conditional generative adversarial networks (ccGAN), which is designed for image generation. The trained model produces high-fidelity XRD spectra for all alloys within the system. **b** The ccGAN consists of a generator and a discriminator. Noise and alloy composition are the inputs of the generator, which generates XRD spectra. The discriminator compares the generated spectra and experimental spectra until the generated spectra cannot be distinguished from the experimental ones.

structure and glass formation. However, characterizing every alloy using XRD is impractical, especially given the vast compositional space defined by the diverse selection of constituent elements in MG development[1,9]. Ideally, the XRD spectrum of any alloy within a system can be accurately predicted, allowing the identification of glass-forming alloys based on the predicted spectra. Unlike material properties, which are typically represented as numerical values and can be readily modeled using ML[20,21,27], XRD spectra encode complex, multidimensional information that cannot be treated as simple numerical data. Although numerical features such as peak positions and peak widths in XRD spectra are critical characteristics, they do not encompass the entirety of relevant information. XRD spectra also embody additional features, such as peak symmetry. These pose a significant challenge in applying ML strategies developed for MG properties to the prediction of XRD spectra.

A possible solution to this challenge is to treat XRD spectra as images. An image is fundamentally composed of a matrix of discrete pixel values, and computational algorithms process these matrices to achieve image generation. In the context of XRD spectra, the angular positions can be interpreted as pixel coordinates, while the corresponding intensities can be considered as the pixel values. This would enable the use of ML models designed for image generation to predict XRD spectra. For example, the continuous conditional generative adversarial network (ccGAN) has demonstrated success in image generation[34]. By representing spectra as images, ccGAN may provide a promising framework for modeling and generating XRD spectra for all alloys within a system. This would substantially enhance the efficiency of identifying glass formation across previously unexplored compositional space.

In the present work, we have accomplished this using a generative deep-learning model based on ccGAN. The trained model is able to generate high-fidelity XRD spectra for uncharacterized alloys. The ability of the ccGAN to generate accurate spectra is validated with datasets from multiple ternary and quaternary alloy systems, as well as data collected from synchrotron facilities and molecular dynamics (MD) simulations. We evaluated the model's accuracy by training it with an increasing number of experimentally measured XRD spectra. It was found that the number of spectra required to effectively train the model is small. To capture the variation in XRD spectra with composition, less than 5% of the experimental spectra are needed for training. This dramatically reduces the number of alloys to be characterized, enabling the exploration of glass formation across vast compositional space.

Results

Strategy and modeling

To generate high-fidelity XRD spectra, we first need experimental XRD spectra for training. The workflow is schematically illustrated in Fig. 1. The XRD spectra were obtained from combinatorial alloy libraries synthesized using magnetron co-sputtering[16]. The highly effective cooling rate during sputtering deposition[35] results in vitrification in a wide composition range[15,17,18], allowing us to systematically analyze the tendency of glass formation. The composition spread of the alloy libraries was analyzed by using energy-dispersive X-ray spectroscopy (EDX). The XRD spectra for training were collected using a laboratory X-ray diffractometer, following previously reported methods[16,17], with 2θ values range from 20° to 65°. Since the combinatorial alloy libraries were fabricated and characterized under consistent conditions, we expect generative ML models of high predictability and generalization can be trained with the collected XRD spectra. To cover a wide glass-forming compositional range, we begin with Zr-Cu-Al as a model alloy system, in which many bulk metallic glasses with a critical casting diameter larger than 1 mm were reported[36,37]. Within the Zr-Cu-Al alloy system, 639 sets of EDX and XRD data were collected. Prior to model training, a preprocessing protocol was implemented: (1) subtraction of the Si substrate signal, (2) restriction of 2θ or q values to the same range, (3) intensity normalization via division by maximum intensity followed by scaling to 100 of arbitrary units, and (4) standardization of data length to 3743 points through linear interpolation.

The generation of XRD spectra was accomplished by first training ccGAN, which includes a generator and a discriminator (Fig. 1b). Unlike a traditional generative adversarial network, the generator of ccGAN incorporates not only random noise as input but also the chemical composition as a continuous conditional input. In addition to direct assessment of the generated XRD spectra, the discriminator facilitates a robust correspondence between compositions and XRD spectra. The full width at half maximum (Δq) and the position (q_1) of the first diffraction peak from both generated and real XRD spectra are further compared to evaluate the ability of the generative model to capture the spectrum characteristics.

Generation of laboratory XRD spectra

The XRD spectral dataset for the Zr-Cu-Al alloy system was randomly split into 80% training and 20% testing sets. To evaluate the performance of the trained ccGAN model, we use mean squared error (MSE) to quantify the difference between experimental XRD spectra and generated ones for the alloys in the

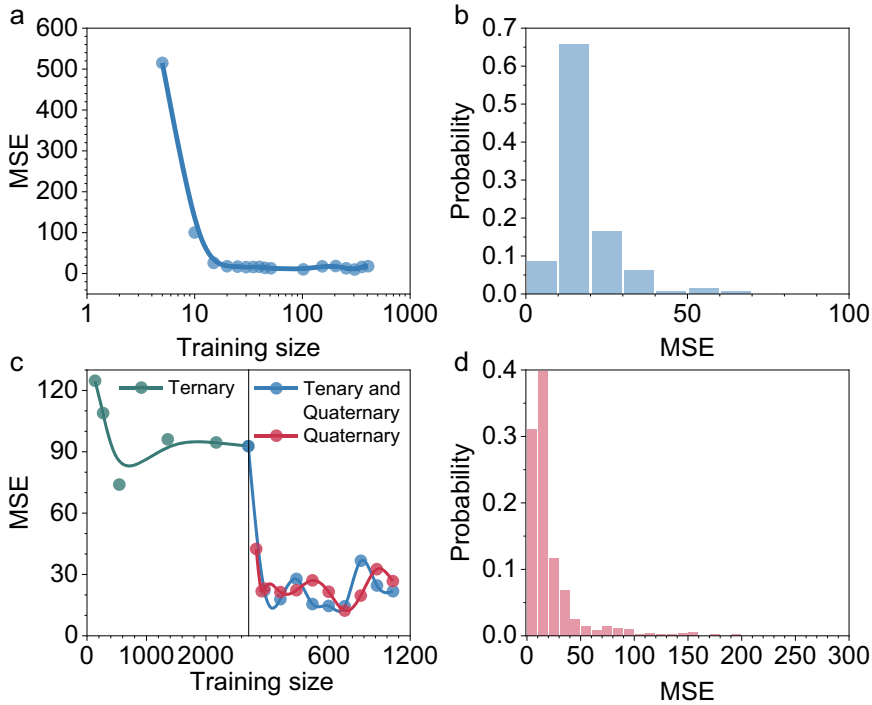

Fig. 2 | Evaluation of the performance of the ccGAN model in generating XRD spectra. **a** Variation of MSE on the Zr-Cu-Al testing set as the training set size increases. **b** Distribution of MSE values on the testing set using the model trained with XRD spectra for 20 randomly selected alloys in the Zr-Cu-Al system. **c** Variation of MSE on Zr-Cu-Ni-Al testing set when the models are trained with spectra only from ternary alloys (red line), only from quaternary alloys (green line), and from both ternary and quaternary alloys (blue line). **d** Distribution of MSE values on the testing set using the model trained with spectra for 100 randomly selected quaternary Zr-Cu-Ni-Al alloys.

testing set. The average MSE over all alloys in the testing set was calculated as the indicator of model performance. Figure 2a illustrates the learning curve of the ccGAN model. It can be seen that MSE dramatically decreases from more than 500 to less than 50 if more XRD spectra are used for training. However, the decrease in MSE slows down when the number of training spectra reaches ~20. MSE as small as 20 is obtained and remains essentially unchanged even if the training dataset increases to more than 100 spectra. In addition to examining the average MSE of the testing set, we also analyze the MSE distribution of the model trained with 20 XRD spectra (Fig. 2b). It turns out that most MSE values for the testing set are below 40. This indicates that we achieved an accurate ccGAN model capable of generating synthetic XRD spectra that are nearly identical to the experimental ones.

It is well known that the compositional space defined by the constituent elements expands exponentially as the number of elements increases[9]. To explore the vast, high-dimensional compositional space, many efforts have been made to predict material properties using low-dimensional information, such as the weighted averages of elemental properties[20,21,30] and binary phase diagrams[23]. However, these predictions are often inaccurate[32]. The success of the ccGAN model in ternary alloys encourages us to apply it to the prediction of XRD spectra for quaternary alloys. As a case study, we take the Zr-Cu-Ni-Al alloy system as an example. We synthesized and characterized alloy libraries of the Zr-Cu-Al, Zr-Cu-Ni, Zr-Ni-Al, and Cu-Ni-Al systems, and collected experimental XRD spectra for 2716 ternary alloys. Additionally, we collected XRD spectra for 1704 quaternary Zr-Cu-Ni-Al alloys. The ccGAN model was trained following three protocols: (1) training with spectra only from ternary alloys (P1), (2) training with spectra only from quaternary alloys (P2), and (3) training with spectra from both ternary and quaternary alloys (P3). For each protocol, various numbers of spectra were used for model training. The learning curves are shown in Fig. 2c. It is noted that the model trained with P1 cannot achieve satisfactory accuracy. Even when more than 2500 spectra (out of 2716) from ternary alloys are used for training, the average MSE remains as high as ~80 on the testing set (see left panel of Fig. 2c). In contrast, the models trained with P2 and P3 exhibit excellent performance, with an average MSE as low as 30. The MSE converges when ~100 XRD spectra are used for training. Further increasing the number of training spectra does not improve the accuracy of the model. Figure 2d shows the distribution of MSE for the testing set, for the model trained with spectra from 100 quaternary Zr-Cu-Ni-Al alloys. It is revealed that only ~10% of the testing set has an MSE greater than 40, indicating the high accuracy of the model. In addition, these results suggest that effective model training requires inclusion of XRD spectra from quaternary alloys, because the model trained with P1 fails to accurately predict XRD spectra for quaternary alloys. The results imply that the model can interpolate within the range of the training data but struggles to extrapolate beyond it. This limitation likely arises from the complex interactions among elements, which make the properties of quaternary alloys different from those of ternary alloys. Furthermore, it is striking that only 5.9% (100 out of 1704) of experimental XRD spectra are needed to effectively train an accurate model.

To visualize the effectiveness of the generative model, we present experimental and model-generated XRD spectra for Zr-Cu-Ni-Al alloys in Fig. 3. The figure displays two examples with MSE below 20 (Fig. 3a, b) and two examples with MSE above 40 (Fig. 3c, d). The MSE increases from 3.57 to 196.80 in Fig. 3a, d. With the increase in MSE, the generated spectra gradually deviate from the experimental ones. This trend suggests that the MSE is an effective indicator of the model's ability to generate high-fidelity XRD spectra. It is remarkable that even when the MSE value is as high as 196.80 (Fig. 3d), the generated spectrum still captures the general trend of the experimental spectrum. Before model training, the only processing of XRD spectra is intensity normalization and the subtraction of signals from the Si substrate. Therefore, the processed spectra contain not only structural information, but also background signals and noise. However, the ccGAN model primarily captures the structural information, rather than the background signals and noise.

Performance on synchrotron XRD and MD simulations

Compared to laboratory X-ray diffractometers, synchrotron X-rays can provide significantly higher signal-to-noise ratios[18]. To validate the generality of the proposed spectrum generation method, we use previously reported synchrotron data for model training, which include 1323 spectra for Co-Ti-Zr, 2582 spectra for Co-V-Zr, and 1306 spectra for Fe-Ti-Nb ternary alloy systems[18]. These spectra are provided in Q-space, spanning a range from 1 to 5 $Å^{-1}$. Within each alloy system, 30% of the spectra are randomly selected as the testing set. The results show that, for all three ternary systems, low MSE values can be achieved with only a few tens of

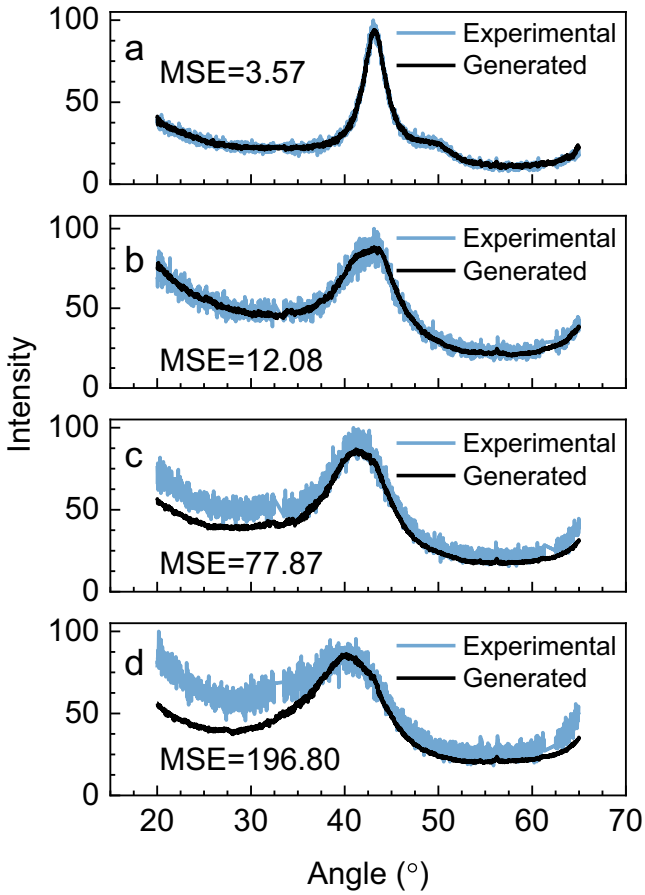

Fig. 3 | Comparison of generated and experimental XRD spectra for Zr-Cu-Ni-Al alloys. **a** The generated spectrum with a MSE of 3.57. **b** The generated spectrum with a MSE of 12.07. **c** The generated spectrum with an MSE of 77.87. **d** The generated spectrum with a MSE of 196.80.

spectra in the training set. Specifically, for the Co-Ti-Zr, Co-V-Zr, and Fe-Ti-Nb systems, the learning curve converges when the training set includes 50 (3.8% of the total spectra), 80 (3.1% of the total spectra), and 60 (4.6% of the total spectra) spectra, respectively. The corresponding MSE values are 45, 13, and 60. Figure 4a–c presents typical examples of the generated spectra for these three alloy systems. Notably, although none of the XRD spectra for the Fe-Ti-Nb system indicate glass formation, the trained model can still generate high-fidelity spectra with sharp peaks indicating the presence of crystalline phases. This suggests that the proposed spectrum generation method is not limited to glass-forming alloy systems but is applicable to crystalline alloys as well. To generate accurate spectra, less than 5% of experimental spectra are required to train the model.

The successful generation of the XRD spectra for both crystalline and glass-forming alloy system implies that the ccGAN model is also applicable to other types of spectra. To verify this, we take radial distribution functions $g(r)$ (RDFs) as an example. We obtain RDFs with MD simulation and train the generative model. For this purpose, we choose the ternary Zr-Cu-Al alloy system because the potential for MD simulation is readily available. Each alloy is cooled from a high temperature of 2000 K to room temperature (300 K) at cooling rates of ~10^{14} K/s. The high cooling rates in the simulations enable glass formation of all the simulated alloys. RDFs of the glassy alloys are calculated at room temperature. We simulated 1176 Zr-Cu-Al alloys at a compositional interval of 2 at% and obtained their RDFs, among which 30% (352 RDFs) are randomly selected as the testing set. From the remaining 70% of the data (824 RDFs), we randomly chose different numbers of RDFs as the training set, thereby enabling the construction of the learning curve. As the training dataset increases to include 40 RDFs (3.4% of the total), an extremely small MSE of 0.0006 is achieved on the testing set (Fig. 4d), and further enlarging the volume of the training set does not improve the accuracy. This indicates that a highly accurate generative ccGAN model can be obtained with less than 4% of the total number of RDFs.

Evaluation on structural characteristics

The positions of the first diffraction peaks q_1 and their full width at half maximum (Δq) are critical parameters that provide essential information about the structural characteristics of an MG[17,33]. We next evaluate the ability of the ccGAN model to reproduce q_1 and Δq. To generate XRD spectra, we use the model trained with data from 100 alloys (see Fig. 2c) in the quaternary Zr-Cu-Ni-Al system. This model is subsequently employed to generate XRD spectra for all 1704 compositions in the system. Notably, both the experimental and the generated XRD spectra are originally represented in 2θ-space. However, since peak positions and peak widths are more physically meaningful when expressed in q-space, all spectra are converted accordingly. We then apply the analysis method reported in ref. 17 to compute q_1 and Δq from both experimental and generated XRD spectra. Figure 5 presents the comparison of the experimental values of Δq and q_1 with those derived from generated spectra. Linear fittings with $y = ax$ show slopes nearly equal to 1. Specifically, a = 0.9907 for Δq and a = 1.0026 for q_1. The coefficients of determination (R^2) for the fittings are 0.9976 and 0.9999, respectively. The excellent fittings confirm the strong linear correlation and the high fidelity of the generated spectra. We reiterate that the training data comprise only 5.9% (100 out of 1704) of spectra from quaternary Zr-Cu-Ni-Al alloys.

Prediction of glass formation within an alloy system

Given that the model can generate XRD spectra and predict Δq and q_1 with high accuracy, we use it to produce spectral data across the entire compositional space of an alloy system and reveal glass formation tendencies via Δq, which is associated with the ability of an alloy to form glass[17]. We first demonstrate this for the ternary Zr-Cu-Al system. To train the generative model, we use experimental XRD spectra from 20 randomly selected alloys. As shown in Fig. 6a, these alloys are sparsely distributed in the compositional space. The large compositional interval between them far exceeds the compositional sensitivity of glass formation[9], preventing any recognition of glass formation tendency from Δq. However, the model trained with spectral data from these 20 alloys enables the generation of XRD spectra at a compositional interval as small as 1 at. % across the system. Figure 6b presents the variation of Δq across the Zr-Cu-Al system using the generated XRD spectra. Since a larger Δq suggests a higher probability for an alloy to form a glass[17], the distribution of Δq shown in Fig. 6b implies that glass formation is most probable in the compositional range of 20–60 at% for Zr, 20–70 at% for Cu, and 0–60 at% for Al. The alloys that exhibit the largest Δq are in the range of 30–60 at% for Zr, 45–65 at% for Cu, and 0–20 at% for Al, well consistent with the range in which bulk MGs were developed[38]. Figure 6c shows the variation of Δq derived from experimental XRD spectra of 639 alloys. Surprisingly, the Δq distribution obtained from the generated XRD spectra is nearly identical to that from the experimental spectra. Although the 20 alloys are sparsely distributed (Fig. 6a), the trained model with their XRD spectra successfully captures detailed glass formation trends across the entire Zr-Cu-Al system (Fig. 6b). We further extend this approach to the quaternary Zr-Cu-Ni-Al alloy system. The model is trained using data from 100 randomly selected quaternary alloys. As in the ternary case, the 100 alloys are far from sufficient to provide detailed insights into glass formation within the Zr-Cu-Ni-Al quaternary system (Fig. 6d). Figure 6e presents the variation of Δq from generated XRD spectra. It can be seen that the revealed tendency is even more comprehensive than that obtained using experimental XRD spectra from 1704 quaternary alloys (Fig. 6f). Moreover, previous attempts[39] to develop bulk glass-forming quaternary alloys were carried out right in the composition range showing larger Δq, further validating that our approach is effective and efficient.

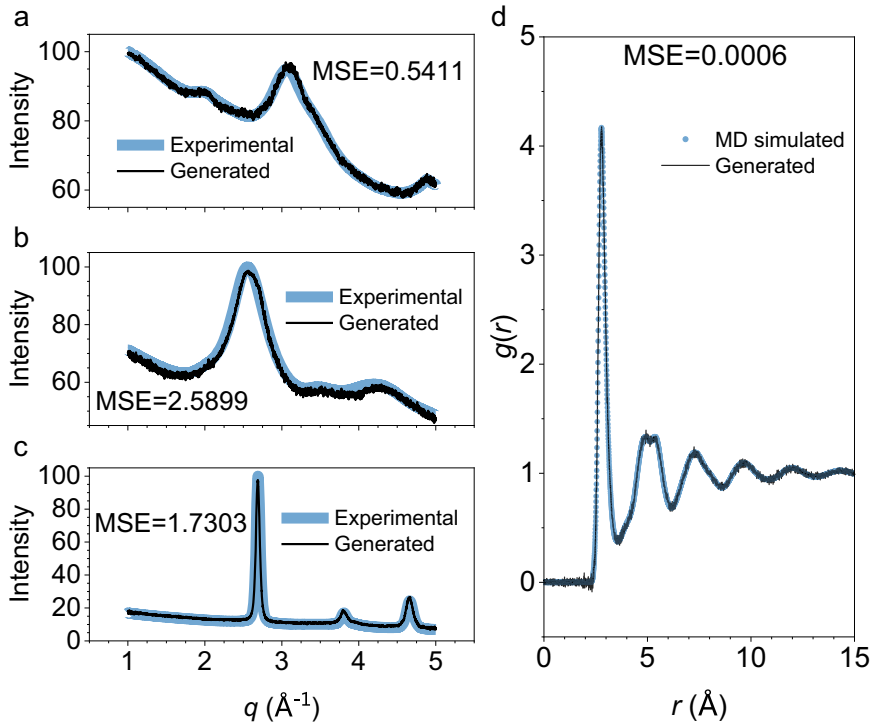

Fig. 4 | Examples of generated synchrotron XRD spectra and radial distribution functions from synchrotron facilities and molecular simulations. **a**–**c** Typical generated synchrotron XRD spectra for Co-Ti-Zr, Co-V-Zr, and Fe-Ti-Nb alloys, with MSE values of 0.5411, 2.5899, and 1.7303, respectively. **d** Typical example of generated RDF based on 40 RDFs from MD simulations. The MSE value is 0.0006.

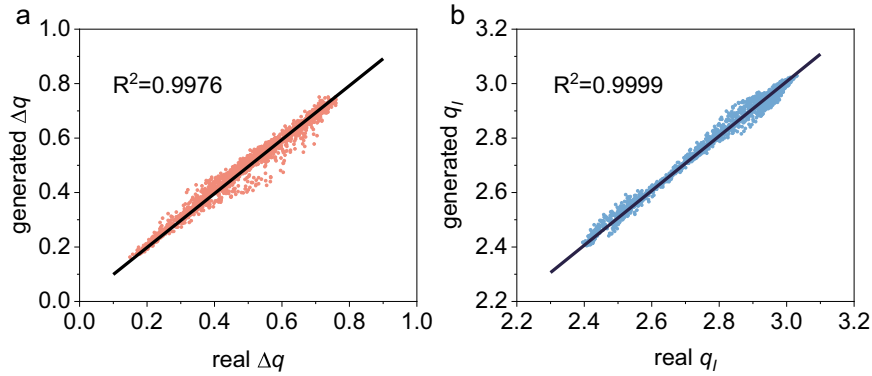

Fig. 5 | Comparison of Δq and q_1 extracted from generated and experimental XRD spectra for 1704 Zr-Cu-Ni-Al alloys. **a** Comparison of Δq. **b** Comparison of q_1. The solid lines in the plots are linear fittings to the data with $y = ax$. The slope is 0.9907 for Δq and 1.0026 for q_1. The coefficient of determination (R^2) is 0.9976 for Δq and 0.9999 for q_1.

Assuming a compositional interval of 1 at. %, one can estimate that there are 4.85×10^3 ternary alloys, 1.56×10^5 quaternary alloys, and 3.76×10^6 quinary alloys per system. The sparse data required to train the accurate generative ccGAN model suggests that far fewer characterizations are needed to reveal glass formation in a multicomponent alloy system than previously assumed. For example, characterizing only a few tens of alloys (~2% of the total) is sufficient for a ternary alloy system, while a few hundred alloys (<1% of the total) are enough for a quaternary system. These indicate that only a tiny fraction of all possible alloys in multicomponent systems are required to be characterized, which falls within the capabilities of current combinatorial methods[15–18]. It is important to emphasize that the quality of generated results depends on the distribution and boundaries of the training data. A uniform distribution of training data across the compositional space is crucial for accurately generating the spectra. If the training data is concentrated in a small compositional range, the model's ability to generalize to new compositional spaces will be compromised.

Discussion

Understanding the atomic structure and establishing the structure-property relationship for MGs has long been pursued. XRD has been and will continue to be the essential tool to confirm glass formation for an alloy. However, laboratory XRD equipment often suffers from inefficient data acquisition. Although synchrotron facilities offer higher efficiency in obtaining XRD spectra, their accessibility remains limited for many research groups. To overcome this limitation, MD simulations have emerged as a tool for probing the atomic packing of MGs, but face constraints in constructing atomic configurations, particularly for alloys of complex chemical stoichiometry.

Unlike previous ML approaches that take material properties as numerical values, we treat the spectra as images. This conceptual shift enables us to use ML models designed for image generation to generate XRD spectra nearly identical to experimental spectra. The approach we introduced substantially facilitates the generation of synthetic spectra for alloys across vast compositional spaces. As we have shown, the model can not only identify the variation of peak position and width with compositions but also can generate the XRD spectra for crystalline alloys. This is particularly valuable for materials design, as it allows for efficiently determining, in complex alloy systems, the regions where composition-induced phase transitions take place. Our approach is also helpful to gain insights into processing-related microstructure evolution. For example, to reveal how the

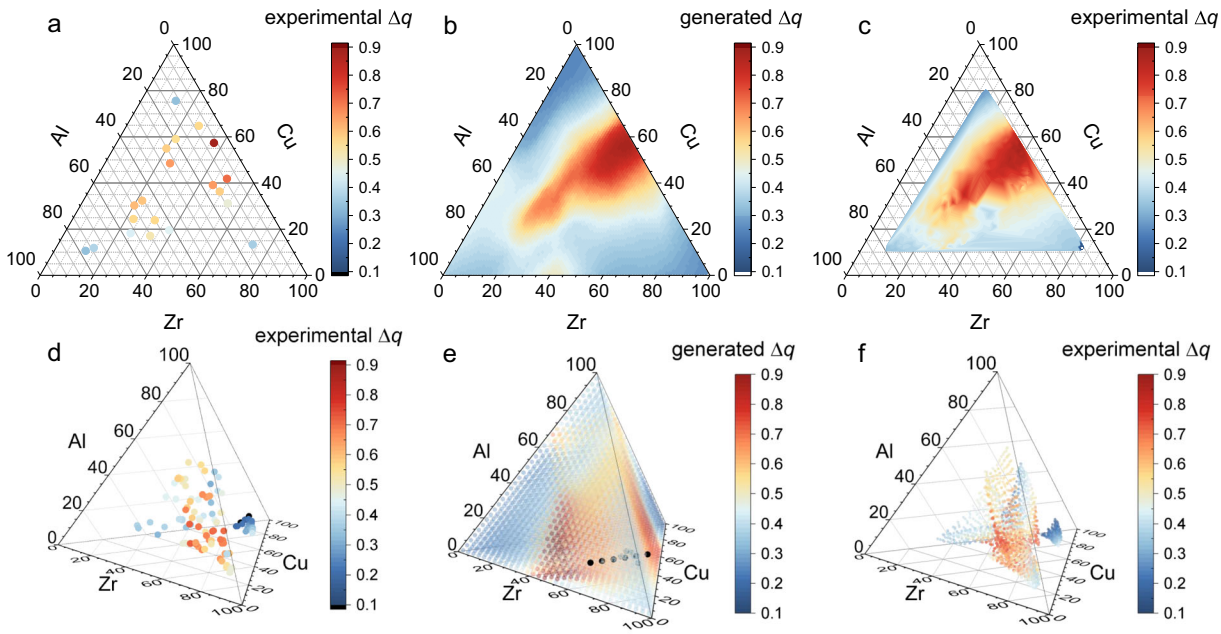

Fig. 6 | Composition-dependent distribution of Δ*q* extracted from experimental and generated XRD spectra. a Δ*q* for 20 randomly selected alloys from the Zr-Cu-Al alloy system for model training. **b** Δ*q* extracted from XRD spectra generated by the model trained with spectra from the 20 alloys shown in (**a**). **c** Δ*q* extracted from XRD spectra by high-throughput experiments covering 639 Zr-Cu-Al alloys. **d** Δ*q* for 100 randomly selected alloys from the Zr-Cu-Ni-Al alloy system. **e** Δ*q* extracted from XRD spectra generated by the model trained with spectra from the 100 alloys shown in (**d**). The black dots in the plot mark the previously reported bulk glass-forming alloys in the system. **f** Δ*q* extracted from XRD spectra by high-throughput experiments covering 1704 Zr-Cu-Ni-Al alloys.

microstructures depend on processing, one only needs to carry out XRD characterizations on materials processed under some conditions, and the rest can be generated by an ML model. This would substantially improve the efficiency and reduce the cost. For simulation of MG structures, our approach simplifies structural modeling because simulations are only needed for alloys of simple stoichiometry, and the structural information for alloys of complex stoichiometry can be generated by the ML model. Here, the simple stoichiometry refers to compositions with elemental contents that can be expressed as integers, such as $Zr_{50}Cu_{30}Ni_{10}Al_{10}$, while the complex stoichiometry refers to chemical compositions with elemental concentrations with decimal precision, such as $Zr_{63.5}Cu_{10.7}Ni_{15.1}Al_{10.7}$. Eliminating the need for complex structural modeling and large-scale atomic system calculations, remarkable computational efficiency with high precision can be achieved with our approach. Furthermore, computational demands can be dramatically reduced because simulations and calculations for an alloy system can be accomplished with a minimal number of configurations.

In summary, by taking a strategic detour that bypasses the complex selection of descriptors, inaccurate material property representation, and insufficient material datasets, we proposed an accurate approach for the identification of glass-forming alloys. We use XRD spectra as an intermediate link between alloy composition and glass formation. Instead of taking the spectra as numerical material properties, we represent them as images, allowing the application of generative machine learning models designed for image generation to the prediction of metallic glasses. Glass formation indicators extracted from the generated XRD spectra capture the details of glass formation tendency across multicomponent alloy systems. The discovery is surprising because it would ultimately allow the development of metallic glasses through data science approaches. The strategy we introduced is not limited to MGs and XRD but is applicable to various materials and spectroscopic techniques. The broad applicability, high efficiency, and accuracy of the proposed approach would accelerate comprehensive exploration of complex material systems within a vast compositional space with diverse characterization methods.

Methods

Alloy library fabrication and characterization

Combinatorial thin-film libraries were fabricated via magnetron co-sputtering deposition (SKY Technology Development Co., Ltd., Chinese Academy of Sciences, TRP450) onto 100-mm-diameter single-side-polished Si wafers, using pure metal sputtering targets with a purity of over 99.95%. For each alloy system, 3–4 experiments with adjusted sputtering power were conducted to cover the majority of the compositional space. Compositions and structures of the alloys were automatically characterized by EDX attached to a Phenom scanning electron microscope and a Malvern Panalytical Empyrean X-ray diffractometer with a Cu Kα radiation source, respectively. XRD spectra for each alloy were processed using programs developed in ref. 17 to calculate the position and full width at half maximum of the first diffraction peak for amorphous phases.

Continuous conditional generative adversarial network (ccGAN)

The architecture of the generator comprises an initial linear layer followed by a transposed convolutional network. The linear layer serves to amalgamate the noise with the input composition. The transposed convolutional network is structured with five transposed convolutional layers, characterized by output channels of [384, 384, 256, 96, 1], kernel sizes of [5, 5, 5, 7, 15], and strides of [3, 3, 3, 3, 5]. Batch normalization and the ReLU activation function were employed in the first four transposed convolutional layers. The output from the final transposed convolutional layer undergoes a sigmoid activation function, which constrains the resultant values within the range of 0–1, subsequently scaled by a factor of 100. The discriminator was constructed with a convolutional network succeeded by two fully connected layers. This convolutional network consists of five convolutional layers, exhibiting output channels of [96, 256, 384, 384, 256], kernel sizes of [11, 5, 3, 3, 3], and strides of [4, 2, 1, 1, 1]. The outputs from the convolutional layers are directed into two fully connected layers. The first fully connected layer features a singular output neuron followed by a sigmoid activation function, while the second fully connected layer is designed to have a number of neurons that corresponds to the size of the composition.

MD simulations

MD simulations with a system size of 40,000 atoms were carried out using LAMMPS software[40]. An embedded atom model potential developed by Sheng et al. was adopted to describe the interatomic interactions[41]. To ensure the comparability of simulation outcomes, an identical MD simulation protocol was employed for each composition. Initially, the alloy systems were relaxed at 2000 K for 20 ps under the NVT ensemble. Then, the ensemble was switched to the NPT during quenching from 2000 K to 300 K at a cooling rate of ~10^{14} K/s, resulting in the formation of amorphous structures. Finally, a 20 ps relaxation was conducted at 300 K. The radial distribution function of the final configuration was calculated using a cutoff distance of 16.0 Å with 3000 bins in the NVE ensemble over 1000 simulation steps, applying time-averaged sampling every 1000 steps to obtain statistically meaningful data.

Data availability

The authors declare that the data supporting the findings of this study are included within the paper. The data are available from the corresponding author upon reasonable request.

Code availability

The code is available from the corresponding author upon reasonable request.

Received: 28 March 2025; Accepted: 23 July 2025;
Published online: 05 August 2025

References

1. Inoue, A. Stabilization of metallic supercooled liquid and bulk amorphous alloys. *Acta Mater.* **48**, 279–306 (2000).
2. Ashby, M. F. & Greer, A. L. Metallic glasses as structural materials. *Scr. Mater.* **54**, 321–326 (2006).
3. Li, F. C. et al. Achieving diamond-like wear in Ta-rich metallic glasses. *Adv. Sci.* **10**, 2301053 (2023).
4. Kumar, G., Tang, H. X. & Schroers, J. Nanomoulding with amorphous metals. *Nature* **457**, 868–872 (2009).
5. Saotome, Y., Itoh, K., Zhang, T. & Inoue, A. Superplastic nanoforming of Pd-based amorphous alloy. *Scr. Mater.* **44**, 1541–1545 (2001).
6. Schroers, J. Bulk metallic glasses. *Phys. Today* **66**, 32–37 (2013).
7. Klement, W., Willens, R. H. & Duwez, P. Non-crystalline structure in solidified gold-silicon alloys. *Nature* **187**, 869–870 (1960).
8. Inoue, A. & Takeuchi, A. Recent development and application products of bulk glassy alloys. *Acta Mater.* **59**, 2243–2267 (2011).
9. Li, Y. L., Zhao, S. F., Liu, Y. H., Gong, P. & Schroers, J. How many bulk metallic glasses are there? *ACS Comb. Sci.* **19**, 687–693 (2017).
10. Chattopadhyay, C., Idury, K. S. N. S., Bhatt, J., Mondal, K. & Murty, B. S. Critical evaluation of glass forming ability criteria. *Mater. Sci. Tech.* **32**, 380–400 (2016).
11. Lu, Z. P. & Liu, C. T. Glass formation criterion for various glass-forming systems. *Phys. Rev. Lett.* **91**, 115505 (2003).
12. Miracle, D. B. A structural model for metallic glasses. *Nat. Mater.* **3**, 697–702 (2004).
13. Guan, P. F., Fujita, T., Hirata, A., Liu, Y. H. & Chen, M. W. Structural origins of the excellent glass-forming ability of PdNiP. *Phys. Rev. Lett.* **108**, 175501 (2012).
14. Cheng, Y. Q. & Ma, E. Atomic-level structure and structure-property relationship in metallic glasses. *Prog. Mater. Sci.* **56**, 379–473 (2011).
15. Ding, S. Y. et al. Combinatorial development of bulk metallic glasses. *Nat. Mater.* **13**, 494–500 (2014).
16. Li, M. X. et al. High-temperature bulk metallic glasses developed by combinatorial methods. *Nature* **569**, 99–103 (2019).
17. Li, M. X. et al. Data-driven discovery of a universal indicator for metallic glass forming ability. *Nat. Mater.* **21**, 165–172 (2022).
18. Ren, F. et al. Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Sci. Adv.* **4**, eaaq1566 (2018).
19. You, D. et al. Electrical resistivity as a descriptor for classification of amorphous versus crystalline phases of alloys. *Acta Mater.* **231**, 117861 (2022).
20. Ward, L. et al. A machine learning approach for engineering bulk metallic glass alloys. *Acta Mater.* **159**, 102–111 (2018).
21. Liu, X. D. et al. Machine learning-based glass formation prediction in multicomponent alloys. *Acta Mater.* **201**, 182–190 (2020).
22. Liu, Z., Chen, C., Zhou, Y. X., Zhang, L. T. & Wang, H. An element-wise machine learning strategy to predict glass-forming range of ternary alloys enabled by comprehensive data. *Scr. Mater.* **229**, 115347 (2023).
23. Sun, Y. T., Bai, H. Y., Li, M. Z. & Wang, W. H. Machine learning approach for prediction and understanding of glass-forming ability. *J. Phys. Chem. Lett.* **8**, 3434–3439 (2017).
24. Cassar, D. R., de Carvalho, A. C. P. L. F. & Zanotto, E. D. Predicting glass transition temperatures using neural networks. *Acta Mater.* **159**, 249–256 (2018).
25. Wang, Q. & Jain, A. A transferable machine-learning framework linking interstice distribution and plastic heterogeneity in metallic glasses. *Nat. Commun.* **10**, 5537 (2019).
26. Fan, Z., Ding, J. & Ma, E. Machine learning bridges local static structure with multiple properties in metallic glasses. *Mater. Today* **40**, 48–62 (2020).
27. Xiong, J., Shi, S. Q. & Zhang, T. Y. A machine-learning approach to predicting and understanding the properties of amorphous metallic alloys. *Mater. Des.* **187**, 108378 (2020).
28. Lu, Z. C. et al. Interpretable machine-learning strategy for soft-magnetic property and thermal stability in Fe-based metallic glasses. *Npj Comput. Mater.* **6**, 187 (2020).
29. Yang, S. Y. et al. Designing Fe-based amorphous alloys with both ultra-high magnetization and ultra-low coercivity through artificial intelligence. *Adv. Funct. Mater.* 2425588 (2025).
30. Zhou, Z. Q., Shang, Y. H., Liu, X. D. & Yang, Y. A generative deep learning framework for inverse design of compositionally complex bulk metallic glasses. *Npj Comput. Mater.* **9**, 15 (2023).
31. Liu, G. N. et al. Machine learning versus human learning in predicting glass-forming ability of metallic glasses. *Acta Mater.* **243**, 118497 (2023).
32. Xie, W. J. L., Sun, M. X., Wang, Y. T., Hu, C. & Liu, L. W. Y.H. A catalogue of metallic glass-forming alloy systems. *Materialia* **39**, 102375 (2025).
33. Ma, D., Stoica, A. D. & Wang, X. L. Power-law scaling and fractal nature of medium-range order in metallic glasses. *Nat. Mater.* **8**, 30–34 (2009).
34. Ding, X., Wang, Y. W., Xu, Z. H., Welch, W. J. & Wang, Z. J. Continuous conditional generative adversarial networks: novel empirical losses and label input mechanisms. *IEEE Trans. Pattern Anal.* **45**, 8143–8158 (2023).
35. Liu, Y. H., Fujita, T., Aji, D. P. B., Matsuura, M. & Chen, M. W. Structural origins of Johari-Goldstein relaxation in a metallic glass. *Nat. Commun.* **5**, 3238 (2014).
36. Cui, X. et al. Centimeter-sized CuZrAl bulk metallic glass with good plasticity and chemical heterogeneity. *Intermetallics* **121**, 106773 (2020).
37. Zhang, Q., Zhang, W., Xie, G. Q. & Inoue, A. Glass-forming ability and mechanical properties of the ternary Cu-Zr-Al and quaternary Cu-Zr-Al-Ag bulk metallic glasses. *Mater. Trans.* **48**, 1626–1630 (2007).
38. Wang, D., Tan, H. & Li, Y. Multiple maxima of GFA in three adjacent eutectics in Zr-Cu-Al alloy system - a metallographic way to pinpoint the best glass forming alloys. *Acta Mater.* **53**, 2969–2979 (2005).
39. Sun, Y. J. et al. Zr-Cu-Ni-Al bulk metallic glasses with superhigh glass-forming ability. *Acta Mater.* **57**, 1290–1299 (2009).

40. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
41. Cheng, Y. Q., Ma, E. & Sheng, H. W. Atomic level structure in multicomponent bulk metallic glass. *Phys. Rev. Lett.* **102**, 245501 (2009).

Acknowledgements

The work was financially supported by the National Natural Science Foundation of China (grant nos. 52331007, 52192602, T2222028, 52471189). The AI-driven experiments, simulations and model training were performed on the robotic AI-Scientist platform of Chinese Academy of Sciences.

Author contributions

Y.H.L. conceived and supervised the research. W.J.X. conducted the experiments with assistance from C.W., M.X.L., and F.C.L. W.J.X. performed MD simulations. W.J.X. and Y.T.S. performed machine learning. W.J.X., Y.T.S., and Y.H.L. wrote the manuscript with input and comments from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Yanhui Liu.

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