

Database of ternary amorphous alloys based on machine learning

Xuhe Gong(巩旭菏), Ran Li(李然), Ruijuan Xiao(肖睿娟), Tao Zhang(张涛), and Hong Li(李泓)
Citation: Chin. Phys. B, 2025, 34 (**1**): 016101. DOI: 10.1088/1674-1056/ad8ec8

What follows is a list of articles you may be interested in

Properties of radiation defects and threshold energy of displacement in zirconium hydride obtained by new deep-learning potential

Xi Wang(王玺), Meng Tang(唐孟), Ming-Xuan Jiang(蒋明璇), Yang-Chun Chen(陈阳春), Zhi-Xiao Liu(刘智骁), and Hui-Qiu Deng(邓辉球)
Chin. Phys. B, 2024, 33 (**7**): 076103. DOI: 10.1088/1674-1056/ad362b

Effect of Y element on atomic structure, glass forming ability, and magnetic properties of FeBC alloy

Jin-Hua Xiao(肖晋桦), Da-Wei Ding(丁大伟), Lin Li(李琳), Yi-Tao Sun(孙奕韬), Mao-Zhi Li(李茂枝), and Wei-Hua Wang(汪卫华)
Chin. Phys. B, 2024, 33 (**7**): 076101. DOI: 10.1088/1674-1056/ad3dd4

Geometries and electronic structures of Zr_nCu(n =2-12) clusters: A joint machine-learning potential density functional theory investigation

Yizhi Wang(王一志), Xiuhua Cui(崔秀花), Jing Liu(刘静), Qun Jing(井群), Haiming Duan(段海明), and Haibin Cao(曹海宾)
Chin. Phys. B, 2024, 33 (**1**): 016109. DOI: 10.1088/1674-1056/acd5c2

Geometries and electronic structures of Zr_nCu(n =2-12) clusters: A joint machine-learning potential density functional theory investigation

Yizhi Wang(王一志), Xiuhua Cui(崔秀花), Jing Liu(刘静), Qun Jing(井群), Haiming Duan(段海明), and Haibin Cao(曹海宾)
Chin. Phys. B, 2024, 33 (**1**): 016109. DOI: 10.1088/1674-1056/acd5c2

Accurate prediction method for the microstructure of amorphous alloys without non-metallic elements

Wei Zhao(赵伟), Jia-Lin Cheng(成家林), Gong Li(李工), and Xin Wang(王辛)
Chin. Phys. B, 2021, 30 (**11**): 116103. DOI: 10.1088/1674-1056/abfd9d

---

DATA PAPER

Database of ternary amorphous alloys based on machine learning

Xuhe Gong(巩旭菏)[1,2], Ran Li(李然)[1,†], Ruijuan Xiao(肖睿娟)[2,‡], Tao Zhang(张涛)[1], and Hong Li(李泓)[2]

[1] *School of Materials Science and Engineering, Key Laboratory of Aerospace Materials and Performance (Ministry of Education), Beihang University, Beijing* 100191, *China*

[2] *Institute of Physics, Chinese Academy of Sciences, Beijing* 100190, *China*

(Received 15 September 2024; revised manuscript received 22 October 2024; accepted manuscript online 5 November 2024)

The unique long-range disordered atomic arrangement inherent in amorphous materials endows them with a range of superior properties, rendering them highly promising for applications in catalysis, medicine, and battery technology, among other fields. Since not all materials can be synthesized into an amorphous structure, the composition design of amorphous materials holds significant importance. Machine learning offers a valuable alternative to traditional "trial-and-error" methods by predicting properties through experimental data, thus providing efficient guidance in material design. In this study, we develop a machine learning workflow to predict the critical casting diameter, glass transition temperature, and Young's modulus for 45 ternary reported amorphous alloy systems. The predicted results have been organized into a database, enabling direct retrieval of predicted values based on compositional information. Furthermore, the applications of high glass forming ability region screening for specified system, multi-property target system screening and high glass forming ability region search through iteration are also demonstrated. By utilizing machine learning predictions, researchers can effectively narrow the experimental scope and expedite the exploration of compositions.

Keywords: amorphous alloys, machine learning, database

PACS: 61.43.Dq, 71.23.Cq

DOI: 10.1088/1674-1056/ad8ec8 **CSTR:** 32038.14.CPB.ad8ec8

1. Introduction

Amorphous materials are known for their unique disordered structures and have a wide range of application scenarios. In the field of electrocatalytic water splitting, for example, amorphous alloys emerge as superior catalysts, outperforming crystalline alloys with the same composition by significantly reducing overpotential.[1,2] Amorphous alloys can also be used as biomedical bone implant materials due to their excellent mechanical properties and corrosion resistance.[3,4] Besides, amorphous solid electrolytes have gained attention for their outstanding electrochemical performance. Amorphous solid electrolytes exhibit excellent ionic conductivity, admirable stability and unique ion transport mechanisms.[5–8] These advantages are linked to the inherent disordered nature of the material to a large extent, making composition design a pivotal area of research focus. In this work, we select amorphous alloys with relatively more experimental data as the research objects for the composition design of disordered structure materials. Research on the prediction of glass forming ability (GFA) can enrich our understanding of the preparation and application of amorphous materials in battery technology. These insights are essential to drive the development of disordered materials for future battery systems.

Through continuous summarization of experimental results, researchers have summarized some empirical laws for GFA to guide the composition design, such as the three principles of Inoue and the eutectic point criterion. Meanwhile, some characteristic indicators such as the normalized glass transition temperature, the width of supercooled liquid region, etc. have also been proposed by some researchers.[9–11] Despite the availability of numerous reports to guide composition design, the development of new compositions still involves extensive experimental "trial-and-error", resulting in high costs as well as low efficiency. In recent years, machine learning has emerged as a powerful tool in composition design, by using known experimental data as training sets to predict unknown properties and guide experiments.[12–16] For amorphous alloys, machine learning models are capable of predicting properties such as characteristic temperature, GFA and mechanical properties, among others. In contrast to traditional empirical rules, machine learning can fully utilize existing data, including experimental results that do not form amorphous structures. This approach enables more precise predictions,[17] especially in scenarios involving complex elemental compositions. Liu *et al.*[18] employed the random forest to construct the classification model of GFA. They demonstrated that the prediction accuracy could be significantly enhanced by grouping based on the classification system proposed by Inoue, and they conducted an analysis of the significance of the GFA descriptive factors within each respective group. Jeon

†Corresponding author. E-mail: liran@buaa.edu.cn

‡Corresponding author. E-mail: rjxiao@aphy.iphy.ac.cn

© 2025 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

et al.[19] utilized the random forest model to predict the glass transition temperature and crystallization temperature of Ni-based amorphous alloys. Subsequently, they designed a novel $Ni_{50.9}Mo_{11.8}Al_{22.3}Nb_{15.0}$ amorphous alloy based on the predictions of the model. Xiong *et al.*[20] predicted the GFA, bulk modulus, and shear modulus through machine learning and discovered a negative correlation between the shear and bulk moduli of bulk amorphous alloys and the average atomic volume of the constituent alloys, and observed that the mixing entropy enhances the shear modulus. Recently, we developed a machine learning model for predicting the critical casting diameter using experimental data. We established an efficient workflow to predict GFA by combining the reduction of data imbalance and the iteration of a small number of experimental data.[21] Through machine learning, the properties of the predicted target can be quickly obtained without experiments, which can significantly expedite composition selection. Consequently, the establishment of a comprehensive prediction database with valuable insights gleaned from machine learning results holds significant value in expediting the composition selection.

In this work, we train machine learning models respectively for three important properties of amorphous alloys, including critical casting diameter (D_c), glass transition temperature (T_g), and Young's modulus (E). The critical casting diameter refers to the maximum diameter that can form amorphous alloys when prepared by the copper mold casting method, which is the most easily obtained and most commonly used property to reflect the GFA. The glass transition temperature is indicative of the thermal stability of the amorphous structure and represents a critical parameter for amorphous alloys in production and application. Additionally, Young's modulus is a pivotal mechanical property of amorphous alloys, and there is theoretical research on modulus-guided composition design, which aims to derive other physical properties based on the modulus.[22] We then use trained models to construct a comprehensive database for common reported ternary bulk amorphous alloy systems and show the application of our database.

2. Data acquisition

In this paper, we develop a machine learning workflow designed to achieve predictions of amorphous alloys' properties through literature data, as illustrated in Fig. 1. First, the training set is derived from extensive literature data across relevant fields. To ensure data quality, we implement a rigorous data cleaning process that includes removing outliers and standardizing formats and units, resulting in a clean and reliable dataset. Next, we extract 29 features from the online matminer database.[23] Analysis of the descriptor linear correlation matrix reveals that the correlation between these features is below 0.85, indicating their suitability for machine learning, as depicted in Fig. S1. During the training process, each property is trained separately using three algorithms: support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGBoost). These three models have good resistance to overfitting and perform well in small datasets. The Hyperopt optimization algorithm,[24] combined with five-fold cross-validation, is employed to minimize the determination coefficient (R^2) and optimize the model parameters. The set of parameters with the biggest R^2 on the test set is selected as the final hyperparameter selection for each model, and the prediction ability of the models compared for different targets is shown in Fig. 2(a). Finally, the SVM is selected to predict the modulus of amorphous alloys, while the XGBoost is selected to predict the glass transition temperature and critical casting diameter. The performance of the various models on the test set is illustrated in Figs. 2(b)–2(d).

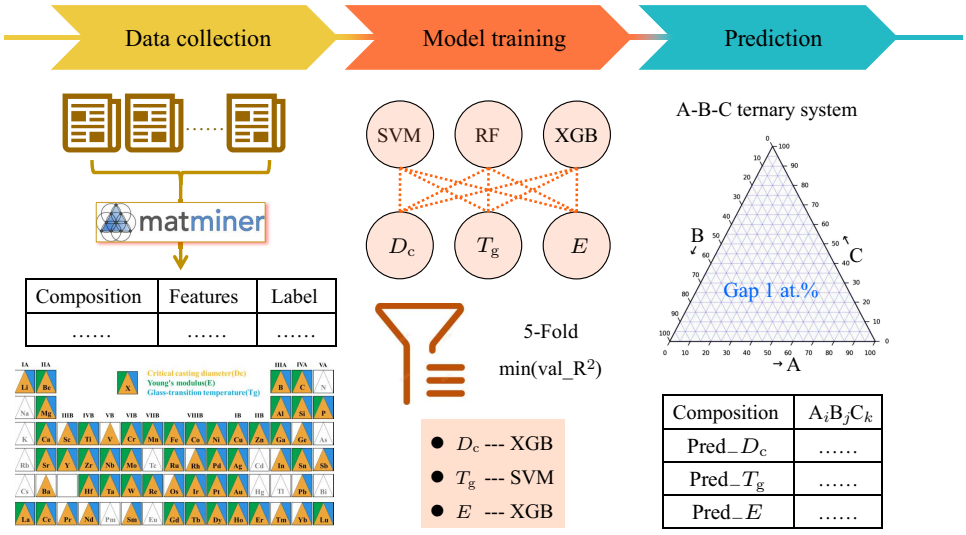


Fig. 1. The machine learning workflow used to obtain the prediction database.

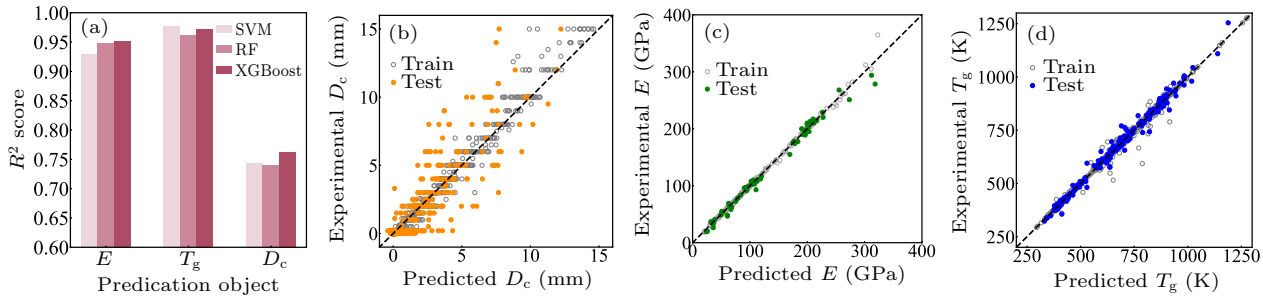


Fig. 2. (a) Prediction effect of different models on the test set. (b)–(d) The final optimized models on D_c, T_g and E prediction, where the horizontal axis represents the predicted value and the vertical axis represents the reported value.

3. Details of database

The database in this study is derived from predictions using the models mentioned above and includes the following 45 alloy systems: Ag–Mg–Ca, Ca–Al–Ag, Ca–Al–Cu, Ca–Al–Zn, Ca–Mg–Al, Ca–Mg–Cu, Co–Nb–B, Co–Ta–B, Ce–Al–Co, Cu–Hf–Ti, Cu–Zr–Al, Cu–Zr–Ti, Fe–P–C, Fe–Nb–B, Ga–Co–Al, Ga–Ni–Al, Hf–Cu–Al, Hf–Ni–Al, Ir–Ni–Ta, La–Al–Co, La–Ni–Al, Mg–Ag–Y, Mg–Cu–Gd, Mg–Cu–Y, Mg–Cu–Yb, Mg–Zn–Ca, Mo–Co–B, Ni–Nb–Ta, Ni–Nb–Zr, Ni–Ta–Sn, Pd–Ag–Si, Pd–Au–Si, Pd–Cu–P, Pd–Cu–Si, Pd–Ni–P, Sr–Mg–Zn, Ta–Ni–Co, Ti–Cu–Be, Ti–Cu–Ni, Ti–Cu–Zr, Ti–Zr–Be, Y–Al–Co, Zr–Al–Co, Zr–Al–Ni, and Zr–Cu–Be. These systems represent commonly reported bulk amorphous alloys and are of significant research interest. Other alloy systems can also be predicted using the same workflow. The distribution of the elements used in these alloy systems is illustrated in Fig. 3(a), with red-highlighted regions indicating their presence in the database. The element name indicates that this element has been reported as one of the elements of amorphous alloys, while those with slash line indicates that no reported amorphous alloys contain this element. This indicates that our database encompasses a broad range of elements known to form amorphous alloys. Figures 3(b)–3(d) show the distribution of the predicted objects across the entire database. It is observed that the predicted critical casting diameter is predominantly concentrated within the region of less than 2 mm, which can be attributed to the fact that only a fraction of the system typically exhibits high GFA. Concurrently, the predicted glass transition temperature and Young's modulus exhibit an approximately normal distribution pattern.

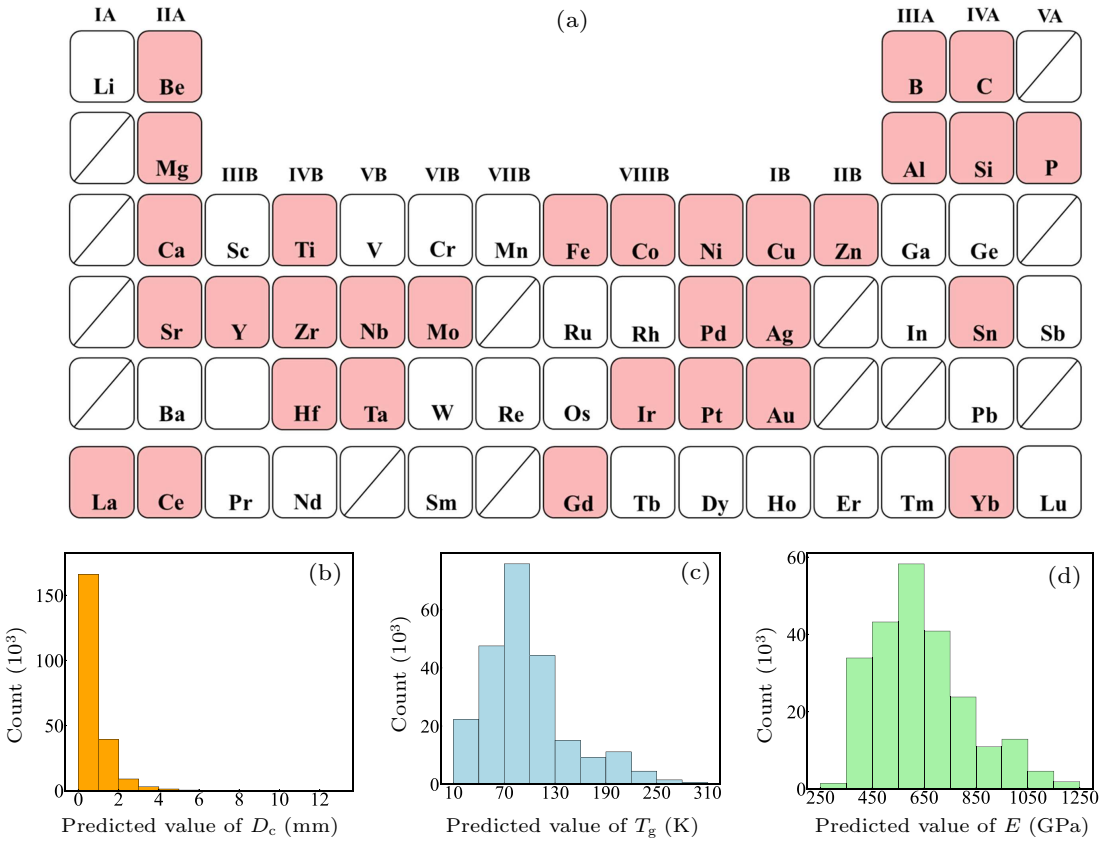


Fig. 3. Statistics of database. (a) Distribution of elements on the periodic table in the predicted database. (b)–(d) Histograms of frequency statistics for D_c, T_g, and E, respectively.

Table 1. Contents for the database.

Content	Type	Comment
system	str	the classification of alloys characterized by a conserved ensemble of fundamental constituent elements
composition	str	the proportion of elements in a specific alloy
pred_D_c (mm)	float	the critical casting diameter of the corresponding composition
pred_T_g (K)	float	the glass-transition temperature of the corresponding composition
pred_E (GPa)	float	the Young modulus of the corresponding composition

The prediction object encompasses all compositions within 45 systems that vary at a step of 1 at.%. The prediction targets are critical casting diameter, glass transition temperature, and Young's modulus for the corresponding compositions. The data are organized into a CSV file, as illustrated in Table 1. The dataset consists of five columns, with the first two columns representing compositional information, indicating the system to which this data belongs and the specific element ratios. The last three columns contain the predicted values, which are the critical casting diameter, glass transition temperature, and Young's modulus, respectively.

4. Application of the database

The database in this study can assist and guide experimenters in designing compositions, reducing experimental costs by combining traditional material science knowledge with prediction results. Specifically, within the same system, the predicted data can be visualized as a ternary heat map to analyze property trends with changes in element composition and narrow down the composition range for experiment. The critical casting diameter of the Ir–Ni–Ta system can be predicted, as illustrated in Fig. 4(a). By considering five atom ratios as the step size, the number of compositions requiring experimental verification can be reduced from hundreds to about twenty. Combined with laser high-throughput and other preparation methods,[25,26] the composition screening can be realized more efficiently. The predicted region shows high GFA contained compositions with high GFA as reported in the experimental study.[27] This correspondence illustrates the generalization ability of the model in the face of unknown systems.

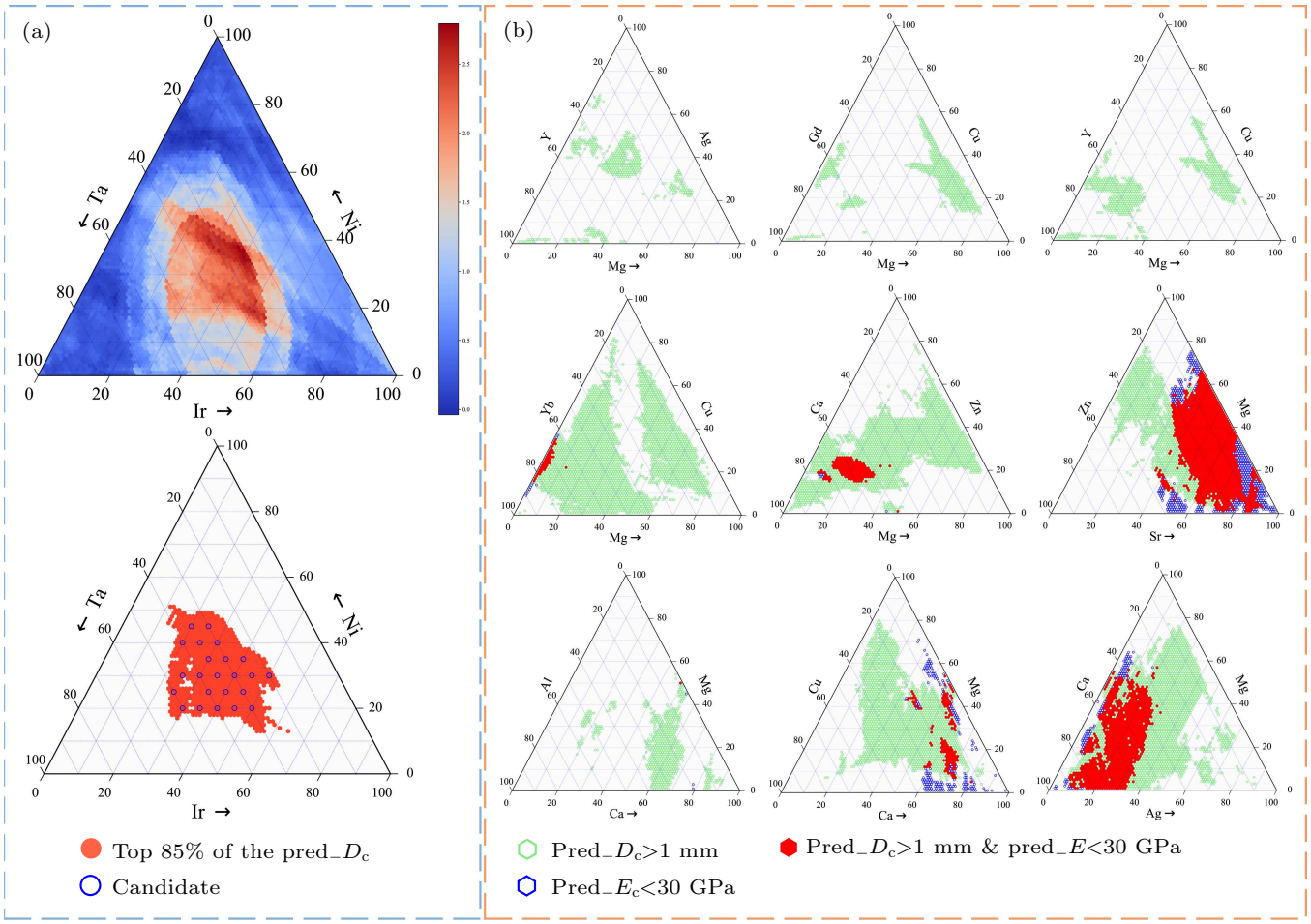


Fig. 4. Examples of database applications. (a) Diagram of searching regions with high GFA in Ir–Ni–Ta system. The top figure is the heat map of the prediction results, and the bottom figure is the screening results. The red part is the top 85% composition region of the prediction results, and the blue hollow circles represent the experimental candidate compositions. (b) Diagram of screening candidate systems for ternary Mg-based biomedical amorphous alloys that meet GFA and T_g requirements. The green area indicates that the predicted critical casting diameter is greater than 1 mm, the blue area indicates that the predicted Young's modulus is less than 30 GPa, and the red area indicates that both of the above are satisfied at the same time.

Moreover, the database can serve as a basis for initial system screening during composition design by predicting numerous potential systems that meet target criteria and setting broad screening conditions to minimize quantity of candidate systems. As shown in Fig. 4(b), it is assumed that the goal is to screen magnesium-based amorphous alloys as bone implant materials while maintaining an elastic modulus similar to human bone. By organizing the prediction data of critical casting diameter and Young's modulus for multiple potential systems in the database, and establishing a set of broad screening criteria such as, the critical casting diameter is greater than 1 mm and the Young's modulus is less than 30 GPa, four candidate systems can be selected through screening. In this case, the Ag–Mg–Ca, Ca–Mg–Cu, Ca–Mg–Cu and Sr–Mg–Zn systems with larger red area are candidate systems. Subsequently, additional conditions like cytotoxicity can be considered for further screening. Prediction can also serve as a part of the target composition discovery process, wherein active learning is established by combining limited experimental data with multiple prediction iterations, which is discussed in Ref. [19].

5. Conclusion

In this paper, we present a comprehensive procedure for establishing a machine learning database of amorphous alloys, encompassing data collection, model training, and prediction. We offer a CSV file containing predicted values for three key physical properties (critical casting diameter, glass transition temperature, and Young's modulus) across 45 systems. Additionally, we add system and element composition tags to facilitate easy retrieval. We demonstrate three typical application directions: composition selection, system selection, and iterative optimization. The methodology and database can be readily extended to encompass various properties of amorphous alloys by simply substituting label values. This database provides efficient and cost-effective design insights for the composition design of amorphous alloys while expediting the experimental composition exploration process.

Data availability statement

The data that support the findings of this study are openly available in Science Data Bank at https://doi.org/10.57760/sciencedb.13096.

Acknowledgements

Project supported by funding from the National Natural Science Foundation of China (Grant Nos. 52172258, 52473227 and 52171150) and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB0500200). The numerical calculations in this study were carried out on both the ORISE Supercomputer and the National Supercomputer Center in Tianjin.

References

[1] Yu X, Gong X, Qiao H, Liu X, Ma C, Xiao R, Li R and Zhang T 2024 *Small Methods* **8** 2400793
[2] Li X, Cai W, Li D S, Xu J, Tao H and Liu B 2021 *Nano Research* **16** 4277
[3] Meagher P, O'cearbhaill E D, Byrne J H and Browne D J 2016 *Adv. Mater.* **28** 5755
[4] Li H F and Zheng Y F 2016 *Acta Biomater* **36** 1
[5] Wu S, Xiao R, Li H and Chen L 2022 *Inorganics* **10** 45
[6] Zhang S, Zhao F, Chen J, Fu J, Luo J, Alahakoon S H, Chang L Y, Feng R, Shakouri M, Liang J, Zhao Y, Li X, He L, Huang Y, Sham T K and Sun X 2023 *Nat. Commun.* **14** 3780
[7] Dai T, Wu S, Lu Y, Yang Y, Liu Y, Chang C, Rong X, Xiao R, Zhao J, Liu Y, Wang W, Chen L and Hu Y S 2023 *Nature Energy* **8** 1221
[8] Hu L, Wang J, Wang K, Gu Z, Xi Z, Li H, Chen F, Wang Y, Li Z and Ma C 2023 *Nat. Commun.* **14** 3807
[9] Lu Z P and Liu C T 2002 *Acta Materialia* **50** 3501
[10] Du X H, Huang J C, Liu C T and Lu Z P 2007 *J. Appl. Phys.* **101** 086108
[11] Lu Z P, Tan H, Li Y and Ng S C 2000 *Scripta Materialia* **42** 667
[12] Shen Z H, Liu H X, Shen Y, Hu J M, Chen L Q and Nan C W 2022 *Interdisciplinary Materials* **1** 175
[13] Liu Y, Niu C, Wang Z, Gan Y, Zhu Y, Sun S and Shen T 2020 *Journal of Materials Science & Technology* **57** 113
[14] Wei J, Chu X, Sun X Y, Xu K, Deng H X, Chen J, Wei Z and Lei M 2019 *InfoMat* **1** 338
[15] Wu Y, Xu B, Sun Y and Guan P 2021 *Chin. Phys. B* **30** 057103
[16] Zhao S, Jiang B, Song K, Liu X, Wang W, Si D, Zhang J, Chen X, Zhou C, Liu P, Chen D, Zhang Z, Ramasamy P, Tang J, Lv W, Prashanth K G, Şopu D and Eckert J 2024 *Materials & Design* **238** 112634
[17] Liu X W, Long Z L, Zhang W and Yang L M 2022 *J. Alloys Compd.* **901** 163606
[18] Liu G, Sohn S, O'hern C S, Gilbert A C and Schroers J 2024 *Acta Materialia* **265** 119590
[19] Jeon J, Kim G, Seo N, Choi H, Kim H J, Lee M H, Lim H K, Son S B and Lee S J 2022 *Journal of Materials Research and Technology* **16** 129
[20] Xiong J, Shi S Q and Zhang T Y 2020 *Materials & Design* **187** 108378
[21] Gong X, Bi J, Liu X, Li R, Xiao R, Zhang T and Li H 2024 *Phys. Rev. Mater.* **8** 055602
[22] Wang W H 2012 *Progress in Materials Science* **57** 487
[23] Ward L, Dunn A, Faghaninia A, Zimmermann N E R, Bajaj S, Wang Q, Montoya J, Chen J, Bystrom K, Dylla M, Chard K, Asta M, Persson K A, Snyder G J, Foster I and Jain A 2018 *Computational Materials Science* **152** 60
[24] James B, Daniel Y, David G C and David D C 2013 *Proceedings of the 30th International Conference on Machine Learning*, June 16–21, 2013, Atlanta, USA, p. 115
[25] Bi J, Liu X, Zhao H, Li R and Zhang T 2024 *Surface and Coatings Technology* **478** 130421
[26] You D, Zhang H, Ganorkar S, Kim T, Schroers J, Vlassak J J and Lee D 2022 *Acta Materialia* **231** 117861
[27] Li M X, Zhao S F, Lu Z, Hirata A, Wen P, Bai H Y, Chen M, Schroers J, Liu Y and Wang W H 2019 *Nature* **569** 99