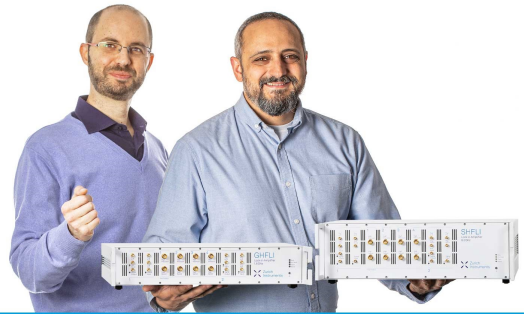
Journal of Applied Physics

Discovery of a new criterion for predicting glass-forming ability based on symbolic regression and artificial neural network

Cite as: J. Appl. Phys. **132**, 125104 (2022); https://doi.org/10.1063/5.0105445
Submitted: 23 June 2022 • Accepted: 25 August 2022 • Published Online: 23 September 2022

Baofeng Tan, Yong-Chao Liang, Qian Chen, et al.



Journal of Applied Physics ARTICLE scitation.org/journal/jap

Discovery of a new criterion for predicting glass-forming ability based on symbolic regression and artificial neural network

Cite as: J. Appl. Phys. **132**, 125104 (2022); doi: 10.1063/5.0105445
Submitted: 23 June 2022 · Accepted: 25 August 2022 ·
Published Online: 23 September 2022

Baofeng Tan, Yong-Chao Liang,[a] Qian Chen, Li Zhang, and Jia-Jun Ma

AFFILIATIONS

Institute of Advanced Optoelectronic Materials and Technology, School of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China

[a]**Author to whom correspondence should be addressed:** 20113248@qq.com

ABSTRACT

Metallic glasses (MGs) are widely used in various fields due to their superior physical properties. Glass-forming ability (GFA) represents the difficulty of forming MGs. Therefore, understanding and establishing the connection between materials characteristics and GFA is a great challenge in MGs research. In this work, to generate a new criterion to characterize GFA, symbolic regression and artificial neural network (ANN) were employed built on 7795 pieces of data. A completely new criterion was proposed and revealed the relationship between three characteristic temperatures (wherein T_g is the glass transition temperature, T_x is the onset crystallization temperature, and T_l is the liquidus temperature) and GFA. The new criterion not only exhibits a higher correlation to the critical casting diameter (D_{max}) than the other 11 reported criteria but also illustrates the importance of high power $(T_x - T_g)/(T_l - T_x)$ in characterizing GFA. Moreover, to test the criterion on unreported data, three models that can, respectively, perform GFA classification, predict D_{max}, and three characteristic temperatures were built through artificial neural networks. Then, 439 new data generated by the ANN model were generated by models applied on Zr–Co–Al–X (X = W, Si, and Ni) alloys. On the testing data, the new criterion shows stronger generalization than other criteria, which proves its reliability and effectiveness.

I. INTRODUCTION

Metallic glass (MG) has been widely studied due to its superior mechanical, physical, and chemical properties. Different from long-range order in crystalline material, the structure of MGs is complex and changeable; therefore, the preparation of new MGs with widespread adoption is confronted with huge challenges.[1–4] Glass-forming ability (GFA) has always been an important problem in MGs research, so understanding and establishing the connection between materials characteristics and GFA is of great significance for the preparation and discovery of MGs. At present, the most important parameters to characterize GFA are the critical cooling rate (R_c) and the critical casting diameter (D_{max}). Usually, a smaller R_c and larger D_{max} imply a better GFA,[5] but the R_c of most alloys is above 10 K/s, and different casting methods have a great influence on D_{max}, which means that they are hard to be precisely measured due to technical limitations. Accordingly, it is extremely important to propose a criterion containing easily measurable parameters to replace R_c and D_{max}. The existing proposed criteria are mainly based on the characteristic temperatures during the solidification process, and the quality performance is judged by studying the Pearson correlation coefficient (r) between the criterion and D_{max}. So far, more than 20 GFA criteria have been found in this way.

At present, some linear GFA criteria have been proposed by traditional methods based on extensive experiences, such as the subcooled liquid region width (ΔT_x) and reduced glass transition temperature (T_{rg}). After that, many dimensionless formulas like $\gamma = T_x/(T_g + T_l)$,[6] $\delta = T_x/(T_l - T_g)$,[7] $\beta = T_x T_g/(T_l + T_x)$,[8] and $\omega = [T_l(T_l + T_x)]/[T_x(T_l - T_x)]$[9] were proposed one after another, which have good correlation with some specific amorphous alloys. This provides new analysis ideas and methods for studying GFA criteria.

With the development of artificial intelligence technology, machine learning as a new research method is gradually introduced

into GFA research. Compared with traditional methods, machine learning has the advantage of a short research period and a low requirement of prior knowledge. For example, a model was developed to predict whether MGs ribbon can be formed by alloy composition.[10] GFA in binary alloys had been studied using a support vector machine (SVM),[11] and a regression neural network model was used to accurately predict the GFA in bulk metallic glass (BMG).[12] Random forest (RF) was used to build a BMG prediction model and three rules regarding the formation of BMG had been proposed.[13] Not only that, machine learning models can be highly predictive for D_{max}, in which r is significantly higher than conventional model predictions.[10,14] All these studies illustrate the reliability and effectiveness of discovering BMG and predicting GFA by machine learning.

Traditional linear GFA criteria need to be considered in the selection of basic quantities, and much domain knowledge is required for researching the model's internal mechanism. The dimensionless formula only focuses on the inherent characteristics of the problem, which is conducive to exploring the objective laws of the model. Therefore, this experiment combined machine learning and dimensionless techniques, and eight dimensionless formulas composed of characteristic temperatures were used to generate a new GFA criterion by the symbolic regression method. Meanwhile, the limited number of data that can be collected makes it difficult to fully test the generalization ability for the new GFA criterion on unreported data. So, this experiment used artificial neural networks (ANNs) to build a complete set of neural network models that can classify GFA and predict D_{max} and characteristic temperatures for any alloys. Then, the new testing data were generated by models applied on Zr–Co–Al–X alloys, which are used to check the generalization ability of the new GFA criterion on new unreported data.

II. COMPUTATIONAL METHODS AND SEQUENCE

A. Data collection

As shown in Fig. 1, the whole experiment was divided into five steps. First, the dataset of this experiment was mainly obtained from the multicomponent alloy provided by Samavatian *et al.*[15] and taking into account the latest experimental data by other researchers.[12,16,17] All data contain 52 different elements and were carefully checked to guarantee their accuracy. The data include the alloy compositions, D_{max}, three characteristic temperatures, and GFA of alloys. The alloys can be classified into three categories by GFA: crystalline alloys (CRAs), ribbon metallic glasses (RMGs), and bulk metallic glasses (BMGs).

B. Data processing

The data were preprocessed in the second step, which consisted of data cleaning and calculation of ANN input features. The data with obvious errors were removed, while the maximum value substitution was applied to the disputed attribute data with the same alloy in different reports. In addition, since the collected data have too little data with D_{max} greater than 5 mm, the log transformation of D_{max} was performed to make it closer to the normal distribution and, thus, improved the generalization ability of the constructed model. After that, 7795 data were filtered as the final data in this experiment, including 4308 RMG data, 1920 CRA data, 1567 BMG data, 1016 D_{max}, and 889 characteristic temperatures, all the data are available in Appendix 1 in the supplementary material.

C. Symbolic regression

Symbolic regression is a machine learning technique that aims to mine the underlying mathematical expressions within the data, which can make connections between two variables without any prior knowledge. It first creates a simple random population of formulas to express the relationship between independent and dependent variables to predict new data and then selects the most suitable individuals from the population for genetic manipulation, which evolves the next generation from the previously generated formulas. The descendant formula undergoes continuous reproduction, mutation, and evolution to eventually move toward the correct answer.

In the third step, to generate a new dimensionless GFA criterion by symbolic regression, 36 dimensionless combinations based on characteristic temperatures proposed by Ren *et al.*[18] were

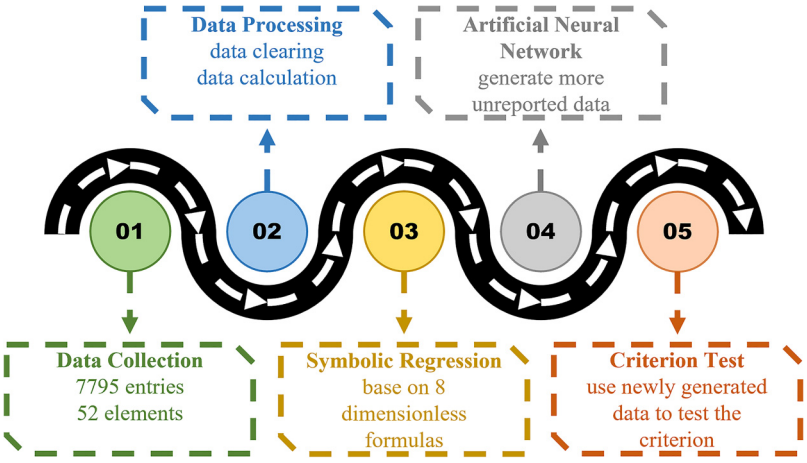


FIG. 1. Schematic of the experimental process.

de-duplicated after mathematical transformation, and finally 23 dimensionless formulas without repetition were obtained, in which the specific formulas are recorded in Appendix 2 in the supplementary material. Since there is a certain randomness in the symbolic regression algorithm, it selected 8 combinations with r greater than 0.4 from 23 combinations, which can reduce the probability of low correlation combinations with D_{max} in the generated criteria. It will also bring down the uncertainty of the algorithm and increase the probability of getting desired results from symbolic regression. The eight selected combinations are $T_x/(T_l - T_x)$, T_x/T_l, $T_g/(T_l - T_x)$, $(T_x + T_g)/(T_l - T_x)$, $T_x/(T_l + T_x)$, $T_x/(T_l + T_g)$, $(T_x - T_g)/(T_l - T_g)$, $(T_x + T_g)/(T_l + T_x)$, and their GFA correlations with D_{max} are shown in Fig. 2.

D. Artificial neural network

Artificial neural network (ANN) refers to a complex network structure formed by a large number of neurons connected, which is an abstraction of the simulation and behavioral mechanism of the human brain.[19] ANN consists of three parts, namely, input layer, hidden layer, and output layer. As the name suggests, the main function of the input layer is to input variables. The hidden layer has multiple neurons, each of which linearly accumulates the input data by its weight, and sends the result to the output layer after calculation by an activation function. The output layer compares the predicted targets with the truthful targets, and the comparison result is returned to the hidden layer through the backpropagation algorithm. It automatically adjusts the weights of the neurons and finally achieves a minimum error between the predicted targets and the truthful targets.

In the fourth step, the new data were generated by the Zr–Co–Al–X alloys. To obtain the GFA classification, D_{max}, and three characteristic temperatures of the new data, three reliable ANN models were constructed. Since the unique properties of each alloy are constituted by its inherent physical, chemical, and thermodynamic properties, as shown in Fig. 3, 25 features[20] that can reflect

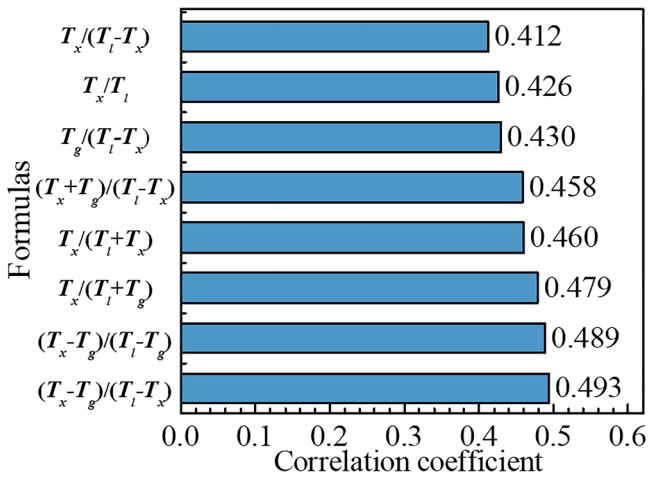


FIG. 2. Histogram of r between formulas and D_{max}.

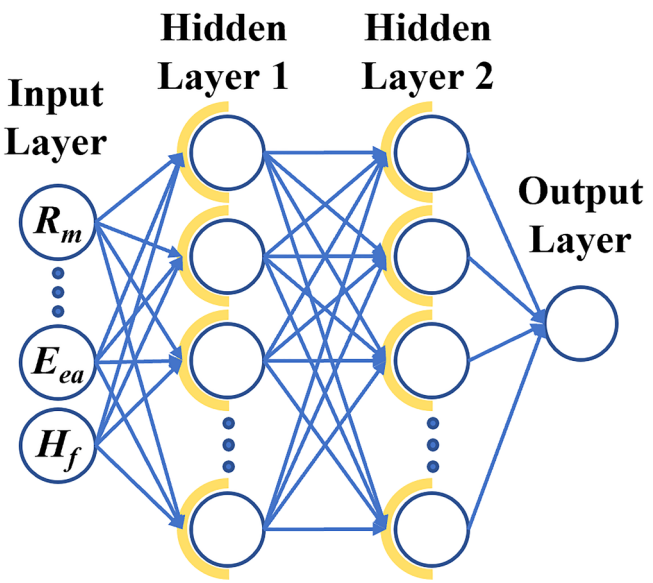


FIG. 3. Schematic of ANN's structure.

the physical, chemical, and thermodynamic properties of alloys were selected as input data of the neural network. These features contain metal radius (R_m), electron affinity energy (E_{ea}), the heat of melting (H_f), specific heat capacity (C_m), etc. All features were scaled to [0, 1] for feature selection and model construction. The collected data were divided into two parts with 70% as training data and 30% as testing data, which means 5456, 711, and 622 data for the training of the GFA prediction model, D_{max} prediction model, and temperatures prediction model, respectively, and 2339, 305, and 267 data for the testing of them separately. The data used for training and testing are carefully selected in the article, covering a variety of possible situations, including 52 elements, and the alloy composition ranges from 2 to multi-element, so it is reliable for training. At the same time, the most suitable hidden layers were selected after many experiments, and the dropout was used before the training process, which enhances the robustness of the model from the perspective of model building.

E. Criterion test

The new GFA criterion was tested on new unreported data at the fifth step. Traditional machine learning divides the data into training and test set and evaluates the merits of the model by observing its performance on the test set. But when the number of the collected data is limited, the test set cannot cover all possible cases and fail to make an objective evaluation. Even if the new GFA criterion performs well on the collected dataset, it is difficult to make it reliable on unreported data. To address this problem, this work generated a new GFA criterion by the collected experimental data and checked its generalization ability to unreported alloys by the new testing data generated in step 4.

III. RESULTS AND DISCUSSION

A. A new GFA criterion

The initial population of symbolic regression was set to 10 000, and the maximum number of iterations was set to 100. The probabilities of crossover and variation were 0.9 and 0.01, respectively, while the penalty coefficient was set to 0.005 to prevent formula inflation. Due to the randomness of symbolic regression, a total of 100 experiments were conducted and an optimal result was selected as shown in Fig. 4. Finally, a new GFA criterion $\theta = \frac{T_g(T_x-T_g)^2}{(T_l-T_x)^3} \times \left(\frac{T_x-T_g}{T_l-T_g} - \frac{T_x}{T_l+T_x}\right)^2$ for characterizing the GFA was obtained by symbolic regression.

Table I shows the Pearson correlation coefficient (r) of 11 reported criteria and the new GFA criterion θ with respect to D_{max}, and the formula for r is as follows:

$$r = \frac{Cov(X, Y)}{\sqrt{Var[X] \times Var[Y]}}, \tag{1}$$

where $Cov(X, Y)$ is the covariance of X and Y, $Var[X]$ is the variance of X, and $Var[Y]$ is the variance of Y. As r is closer to 1, the connection will be closer between the criterion and the D_{max}, which means the better criterion for predicting GFA. The largest r among the 11 reported GFA criteria is G_p, which reaches 0.5476, followed by χ reaching 0.5436. The new GFA criterion θ is higher than all of them, which reaches 0.5587. It fully indicates that the new GFA criterion θ has a stronger performance in predicting GFA.

It is worth noting that the three GFA criterions χ, G_p, and θ with the largest r all contain $(T_x - T_g)/(T_l - T_x)$ in the formulas, while the r of criteria without this combination are all less than 0.5. It is well known that crystallization is a competing process for glass formation and $T_x - T_g$ represents the resistance to crystallization during amorphous formation,[6,28] which can be used to independently represent GFA. $T_l - T_x$ is the same as T_x/T_l in physical significance, representing the liquid phase stability in the crystallization process of supercooled liquid[6] and can also be individually denoted by the strength of GFA. Therefore, there is a necessary connection between $(T_x - T_g)/(T_l - T_x)$ and GFA, which explains why the r of χ, G_p, and θ are much larger than other criteria.

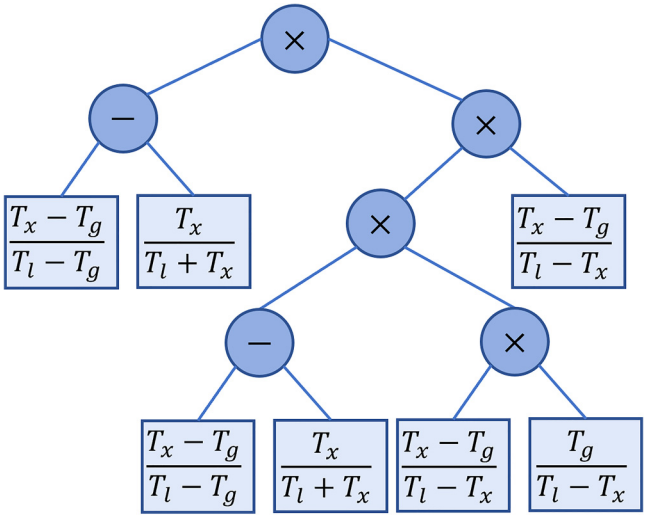


FIG. 4. The result of symbolic regression.

TABLE I. GFA criteria and their Pearson correlation coefficient (r).

No.	GFA parameter	Formula	$r - D_{max}$
1	T_{rg}[21]	T_g/T_l	0.2622
2	γ[6]	$T_x/(T_g + T_l)$	0.4600
3	β_1[22]	$T_x/T_g + T_g/T_l$	0.4691
4	β[23]	$T_g/T_x - T_g/(1.3T_l)$	0.4739
5	φ[24]	$(T_g/T_l) \times [(T_x - T_g)/T_g]^{0.143}$	0.4758
6	β_2[8]	$T_xT_g/(T_l - T_x)^2$	0.4877
7	ΔT_{rg}[25]	$(T_x - T_g)/(T_l - T_g)$	0.4890
8	γ_c[26]	$(3T_x - 2T_g)/T_l$	0.4899
9	ω[9]	$T_l(T_l + T_x)/[T_x(T_l - T_x)]$	0.4927
10	χ[5]	$(T_x - T_g)/(T_l - T_x) \times [T_x/(T_l - T_x)]^{1.47}$	0.5436
11	G_p[27]	$T_g(T_x - T_g)/(T_l - T_x)^2$	0.5476
12	θ (this work)	$T_g(T_x - T_g)^2/(T_l - T_x)^3 \times [(T_x - T_g)/(T_l - T_g) - T_x/(T_l + T_x)]^2$	0.5587

Figures 5(a)–5(c) show the linear fit curves between χ, G_p, θ, and D_{max}, respectively, where the red broad band represents the confidence interval, and the narrower interval indicates the smaller prediction error. All three GFA criteria are closely associated with D_{max}, in which $(T_x - T_g)/(T_l - T_x)$ plays an important role. The new GFA criterion θ is most evenly dispersed, which shows a more pronounced distribution pattern than χ and G_p. The narrower confidence interval indicates a better prediction, which is associated with the higher power of $(T_x - T_g)/(T_l - T_x)$. Not only that, θ also outperforms χ and G_p on large D_{max}, exhibiting stronger generalization ability, which is mainly attributed to the $\left(\frac{T_x-T_g}{T_l-T_g} - \frac{T_x}{T_l+T_x}\right)^2$ in θ.

B. Analysis of new GFA criterion

It is necessary to explain the relationship between θ and GFA from the physical material point of view. The new GFA criterion can be broken down as follows:

$$\theta \propto \frac{T_g}{T_l - T_x}, \tag{2}$$

$$\theta \propto \left(\frac{T_x - T_g}{T_l - T_x}\right)^2 \propto \frac{T_x - T_g}{T_l - T_x}, \tag{3}$$

$$\theta \propto \left(\frac{T_x - T_g}{T_l - T_g} - \frac{T_x}{T_l + T_x}\right)^2 \propto \left(\frac{T_x - T_g}{T_l - T_g} - \frac{T_x}{T_l + T_x}\right). \tag{4}$$

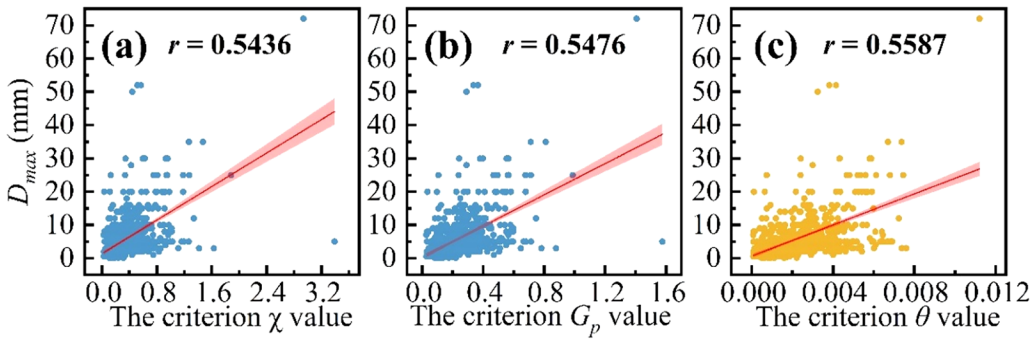


FIG. 5. The correlation plots of (a) χ, (b) G_p, and (c) θ.

According to the traditional kinetic point of view, GFA refers to the ability of an alloy not to produce or to produce dilute crystallization below T_g, so the key to improving GFA is the decrease in its nucleation and growth rate. The nucleation rate (I) and growth rate (U) can be expressed by Eqs. (5) and (6)[29] separately,

$$I = \frac{10^{30}}{\eta}\exp\left[\frac{-16\pi}{3}\times\frac{\alpha^3\Delta S_f T^2}{R(T_l-T)^2}\right], \tag{5}$$

$$U = \frac{10^2}{\eta}\times\left[1-\exp\left(-\frac{\Delta S_f(T_l-T)}{RT}\right)\right]. \tag{6}$$

Here, T is the system temperature. η, α, ΔS_f, and R represent viscosity, a dimensionless parameter related to the liquid–solid interface energy, entropy of melting, and gas constant, respectively. The nucleation and growth rate of the crystal usually depend on the difference between T_x and T_l. Substituting T_x into Eqs. (5) and (6), the following relationships are obtained:

$$I \propto \frac{(T_l-T_x)^2}{T^2}, \tag{7}$$

$$U \propto \frac{T_l-T_x}{T}. \tag{8}$$

As smaller I and U represent stronger GFA, there is

$$GFA \propto \frac{T}{T_l-T_x}, \tag{9}$$

in this context, T represents a specific temperature in the system and T_g is chosen to replace T to obtain

$$GFA \propto \frac{T_g}{T_l-T_x}. \tag{10}$$

Concerning the relationship between GFA and $(T_x-T_g)/(T_l-T_x)$, Tripathi *et al.*[27] verified that supercooled liquids do not crystallize between T_g and T_x, but tend to crystallize between T_x and T_l by Time–Temperature–Transformation (TTT) plots. Therefore, T_x-T_g should be maximized and T_l-T_x should be minimized in the ideal expression for GFA,

$$GFA \propto \frac{T_x-T_g}{T_l-T_x}. \tag{11}$$

The direct proof for Eq. (4) is more difficult and requires a transformation as

$$\theta \propto \left(\frac{1}{\frac{T_l-T_x}{T_x-T_g}+1}-\frac{1}{\frac{T_l}{T_x}+1}\right). \tag{12}$$

Equation (11) shows that $\frac{1}{\frac{T_l-T_x}{T_x-T_g}+1}$ is positively correlated with GFA. When $\frac{1}{\frac{T_l}{T_x}+1}$ is larger, T_l-T_x is smaller, and it can be seen from Eq. (9) that the smaller T_l-T_x represents larger GFA, so GFA and $\frac{1}{\frac{T_l}{T_x}+1}$ are equally positively correlated. Since both are positively correlated with GFA, if they are directly subtracted, the relationship with GFA cannot be analyzed, so it is necessary to subtract $\frac{T_l-T_x}{T_x-T_g}$ from $\frac{T_l}{T_x}$ to compare their growth rate,

$$\frac{T_l-T_x}{T_x-T_g}-\frac{T_l}{T_x}=\frac{T_lT_g-T_x^2}{T_x(T_x-T_g)}. \tag{13}$$

Through the analysis of Eq. (11) and the TTT curve plot in Fig. 6, it can be known that when the GFA becomes larger, T_l-T_x will decrease and T_x-T_g will increase, which means T_l will move closer to T_x and T_x will move away from T_g, so it can obtain that $T_lT_g < T2\,x$. Because the denominator of Eq. (13) is always positive, and the $T_lT_g-T_x^2$ decreases with the increase in GFA, $\frac{T_l-T_x}{T_x-T_g}$ is getting less than $\frac{T_l}{T_x}$.

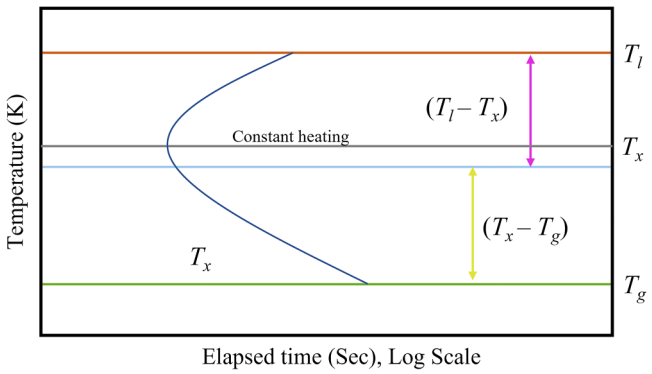


FIG. 6. A schematic representation of the TTT curve showing metallic glass formation.

Eventually, Eq. (12) becomes greater as GFA becomes larger, so it can be proved that

$$GFA \propto \left(\frac{1}{\frac{T_l - T_x}{T_x - T_g} + 1} - \frac{1}{\frac{T_l}{T_x} + 1}\right) \propto \left(\frac{T_x - T_g}{T_l - T_g} - \frac{T_x}{T_l + T_x}\right). \tag{14}$$

Finally, the essence of the new criterion θ is revealed through Eqs. (10), (11), and (14), and Fig. 6. From the material physical point of view, the onset crystallization temperature will become larger with the growth of θ, which substantially increases the width of the subcooled liquid interval, thus increasing the GFA. This is equally well explained why there is a good correlation between θ and GFA, which also theoretically proves its reliability.

It is interesting that, if $\left(\frac{T_x - T_g}{T_l - T_g} - \frac{T_x}{T_l + T_x}\right)^2$ is removed from the formula, the new criterion θ will become $\frac{T_g(T_x - T_g)^2}{(T_l - T_x)^3}$, and the distinction between θ and G_p $\left(\frac{T_g(T_x - T_g)}{(T_l - T_x)^2}\right)$ is only the powers of $(T_x - T_g)$ and $(T_l - T_x)$. Ideally, when GFA is gradually improved, T_x is closer to T_l, and T_g is farther away from T_x, resulting in $T_x - T_g$ gradually larger than $T_l - T_x$. But the different powers in G_p balance the ratio change, which make its overall growth become flat; hence, G_p can outperform on the dataset. However, in practical preparation, the three characteristic temperatures change limitedly as GFA increases. When the powers of $(T_x - T_g)$ and $(T_l - T_x)$ become larger, the size of the power will dominate the change of the formula, thus making $\frac{T_g(T_x - T_g)^2}{(T_l - T_x)^3}$ change rapidly, that is why θ needs other restrictions.

As the powers of $(T_x - T_g)$ and $(T_l - T_x)$ are raised, a restriction factor $\left(\frac{T_x - T_g}{T_l - T_g} - \frac{T_x}{T_l + T_x}\right)^2$ is added to θ. According to the analysis of Eq. (14), $\left(\frac{T_x - T_g}{T_l - T_g} - \frac{T_x}{T_l + T_x}\right)^2$ is proportional to GFA, and it can improve the correlation between θ and GFA. But it can be found that larger $(T_x - T_g)/(T_l - T_x)$ will simultaneously make $\frac{T_x - T_g}{T_l - T_g}$ and $\frac{T_x}{T_l + T_x}$ become bigger, they will grow together and restrict each other, which can reduce the growth rate of $\left(\frac{T_x - T_g}{T_l - T_g} - \frac{T_x}{T_l + T_x}\right)^2$ and make it much slower than $\frac{T_g(T_x - T_g)^2}{(T_l - T_x)^3}$. This also explains why $\left(\frac{T_x - T_g}{T_l - T_g} - \frac{T_x}{T_l + T_x}\right)^2$ can limit the growth rate of $\frac{T_g(T_x - T_g)^2}{(T_l - T_x)^3}$ and enhance the formula expressiveness. It is worth mentioning that both G_p

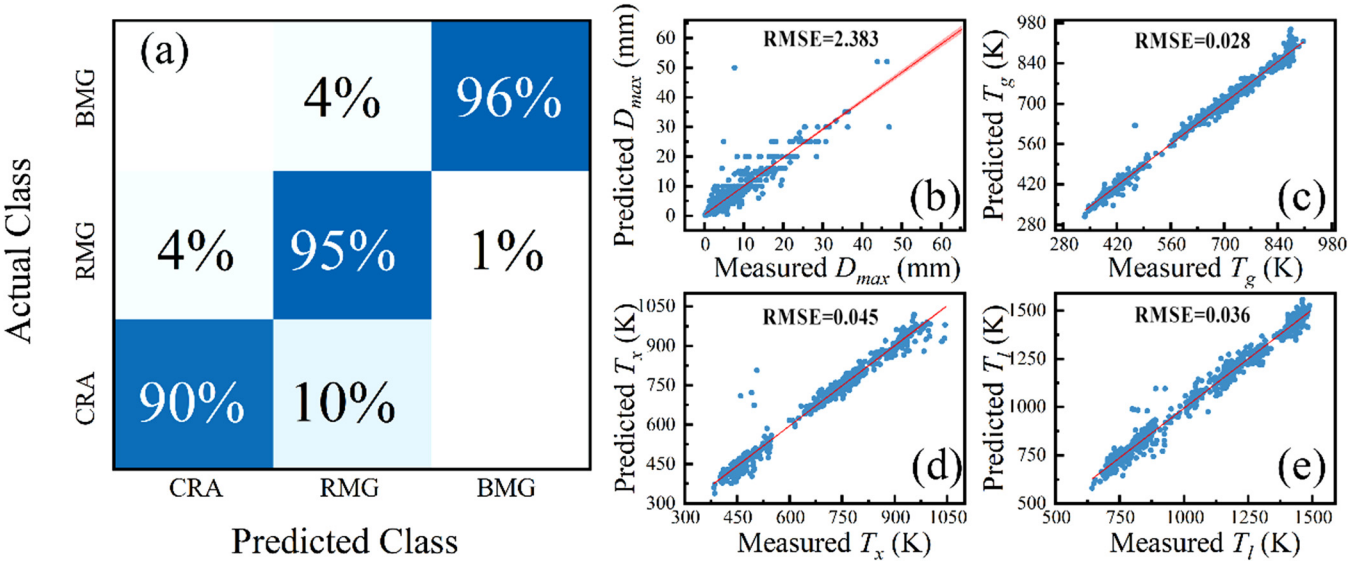


FIG. 7. (a) The confusion matrix of the classification model; the efficiency of regression models for (b) D_{max} and (c)–(e) characteristic temperatures.

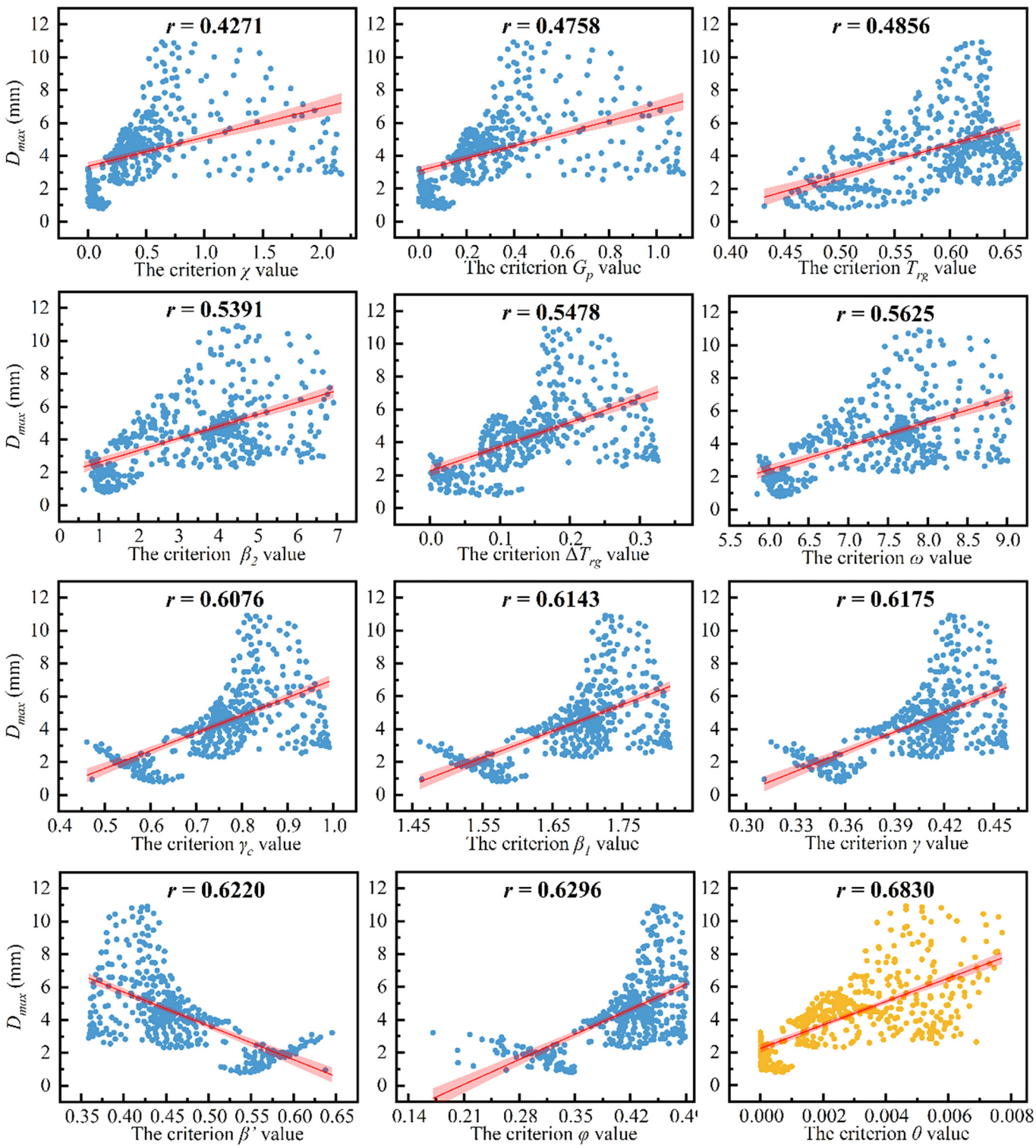


FIG. 8. The correlation plots for χ, G_p, T_{rg}, β_2, ΔT_{rg}, ω, γ_c, β_1, γ, β', φ, and θ with respect to D_{max} on new testing data.

and the new GFA criterion θ explore the predictive effect of the higher power of $\frac{T_x-T_g}{T_l-T_x}$ on GFA, which illustrates that the higher power of $\frac{T_x-T_g}{T_l-T_x}$ does serve as an important parameter for characterizing GFA, but the growth trend of the formula should be considered as increasing the power.

C. New testing data by ANN

In general, generalization ability corresponds to the adaptability of the GFA criterion to new alloy data. But, due to the poor insights into the amorphous structure and the limitation of realistic preparation, only a limited amount of alloy data can be accessed by most experiments, resulting in a scarcity of new data, and the consequent models or criteria are difficult to be applicable on new alloys with unreported data. Therefore, this work generated a series of new testing data by ANN models, so as to verify the generalization ability of the new GFA criterion θ, which is of guiding significance for θ.

Figure 7(a) demonstrates the confusion matrix of the GFA classification model, and it has 90% accuracy for CRA, 95% accuracy for RMG, and the highest accuracy of 96% for BMG. The classification model performs well on the dataset, especially for the BMG data. It shows the predicted D_{max}, T_g, T_x, and T_l comparing the real values in Figs. 7(b)–7(e), in which the root mean squared error (RMSE) is only 2.383 for D_{max} prediction, 0.028 for T_g prediction, 0.045 for T_x prediction, and 0.036 for T_l prediction. For the prediction of these four features, the models show significant convergence and good prediction performances. Obviously, there are deviation points where the fitted curve of the D_{max} regression model in Fig. 7(b) differs significantly from the predicted values. After data checking, the alloys of the deviation points were found to be $Zr_{46}Cu_{31.64}Ag_{8.36}Al_8Be_6$ ($D_{max} = 35$ mm), $Cu_{36}Zr_{48}Ag_8Al_8$ ($D_{max} = 25$ mm), and $(Ti_{41}Zr_{25}Be_{26}Ni_8)_{98}Cu_2$ ($D_{max} = 30$ mm), separately. These deviations are Zr- and Cu-rich alloys with large D_{max}, which can be explained by two possible reasons for this phenomenon. First, there is a shortage of data with D_{max} greater than 20 mm, thus leading to inaccurate predictions of D_{max} for such alloys. Another explanation is that there is too little Zr- and Cu-rich data in the collected dataset, and it also directly affects the prediction effect of the model. The deviation points of three characteristic temperature predictions correspond to $Ag_{30.8}Mg_{23.1}Ca_{30.7}Cu_{15.4}$ and $Zr_{49}Ti_{17}Ni_{20}Cu_{14}$ alloys, which is more likely that the raw data of two alloys are inherently unreliable. A small number of other deviation points exist in the model, but they are within the normal error range.

In recent years, Zr–Co–Al alloys are considered ideal MGs with good GFA and larger D_{max}.[30–32] Meanwhile, considering the effect of trace elements on GFA, $(Zr_iCo_jAl_k)_{95}X_5$ (X = W, Si, and Ni) alloys were chosen to generate the new testing data, in which Zr element grows at 2% in the range of $0 \le i \le 100\%$ and Co element also grows at 2% in the range of $0 \le j \le 100\%$ while ensuring that $i + j + k = 100\%$.

The values of input features were calculated by the compositions of Zr–Co–Al–X alloys, and they were entered into the classification model as input data, then 516 BMGs were identified from the 3978 alloys. The data of 516 BMGs were input into the regression models to predict their D_{max} and characteristic temperatures. To make the data more reliable, data that do not satisfy the size relationship ($T_g < T_x < T_l$) among three characteristic temperatures were removed, and those with $D_{max} > 10$ mm were deleted due to the weakness of the regression model for large D_{max} prediction. After processing, 439 new testing data with confidence were finally gained, and they can be checked in Appendix 3 in the supplementary material.

D. Test of generalization ability

The new GFA criterion θ and 11 other reported criteria were applied to the data generated by the ANN model and the linear fitting curves are drawn individually as shown in Fig. 8. It can be viewed that the r of θ reaches 0.683 and the maximum value of r among the 11 reported criteria is 0.6296. It also realized that χ and G_p perform generally on the new testing data, in which r are only 0.4271 and 0.4758, indicating their poor generalization ability. According to the previous analysis in Sec. III B, the higher powers of $(T_x - T_g)$ and $(T_l - T_x)$ make θ have a stronger generalization ability than χ and G_p. The higher powers of $(T_x - T_g)/(T_l - T_g)$ can characterize GFA and improve the expressiveness of GFA criteria on unreported data.

After the examination of the new testing data, it is reasonable to believe that the new GFA criterion θ is strongly associated with GFA and can predict it better.

IV. CONCLUSIONS

In this work, a new GFA criterion $\theta = \frac{T_g(T_x-T_g)^2}{(T_l-T_x)^3} \times \left(\frac{T_x-T_g}{T_l-T_g} - \frac{T_x}{T_l+T_x}\right)^2$ for predicting GFA was proposed through symbolic regression. The $r = 0.5587$ between θ and D_{max} is significantly better than the other 11 reported criteria. Based on the dimensionless formula, θ omits the choice of fundamental quantities and concentrates on objective patterns. Then, θ was successfully interpreted from a physical material perspective, and it further revealed that higher powers of $(T_x-T_g)/(T_l-T_g)$ have a significant function in characterizing GFA and enhancing the generalization ability. Moreover, to observe the performance of θ on new alloys with unreported data, three neural network models that can classify GFA and predict D_{max} and characteristic temperatures were constructed due to the limited amount of collected data. The models all converged significantly and showed good classification and prediction performance. Additionally, the Zr–Co–Al–X alloys were applied to the neural network model, and 439 new unreported testing data were predicted to validate the generalization ability of θ on new alloys. Finally, the results proved that θ has a stronger ability to predict GFA, and it is equally reliable on unreported data.

SUPPLEMENTARY MATERIAL

See the supplementary material for the preprocessed data, 23 dimensionless formulas, and the new testing data generated by ANN.

ACKNOWLEDGMENTS

This work has been supported by the National Natural Science Foundation of China (NNSFC) (Grant Nos. 11964005,

11963003, 12004053, and 62163006), the Fostering Project of Guizhou University (Grant Nos. [2020]33 and [2020]76), Industry and Education Combination Innovation Platform of Intelligent Manufacturing and Graduate Joint Training Base at Guizhou University (Grant No. 2020-520000-83-01-324061). The authors acknowledge the financial support of the GHfund B (20210702) and Basic Research Program of Guizhou Province (Grant Nos. ZK [2022] 042 and ZK[2022] 143).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Baofeng Tan: Conceptualization (equal); Data curation (equal); Formal analysis (equal); Visualization (equal). **Yong-Chao Liang:** Formal analysis (equal); Methodology (equal). **Qian Chen:** Validation (equal). **Li Zhang:** Data curation (equal); Formal analysis (equal). **Jia-Jun Ma:** Data curation (equal).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its supplementary material.

REFERENCES

[1]Y. Sun, A. Concustell, and A. L. Greer, Nat. Rev. Mater. **1**, 16039 (2016).
[2]W. H. Wang, Prog. Mater. Sci. **106**, 100561 (2019).
[3]S. Majid, G. Reza, A.A. Ahmad, and M. Shamsoddin, Mater. Sci. Eng. A **753**, 218 (2019).
[4]J. Schroers, Adv. Mater. **22**, 1566 (2010).
[5]Z. Long, W. Liu, M. Zhong, Y. Zhang, M. Zhao, G. Liao, and Z. Chen, J. Therm. Anal. Calorim. **132**, 1645 (2018).
[6]Z. P. Lu and C. T. Liu, Acta Mater. **50**, 3501 (2002).
[7]Q. Chen, J. Shen, D. Zhang, H. Fan, J. Sun, and D. G. McCartney, Mater. Sci. Eng., A **433**, 155 (2006).
[8]Z.-Z. Yuan, S.-L. Bao, Y. Lu, D.-P. Zhang, and L. Yao, J. Alloys Compd. **459**, 251 (2008).
[9]X. Ji and Y. Pan, Trans. Nonferr. Met. Soc. China **19**, 1271 (2009).
[10]L. Ward, S. C. O'Keeffe, J. Stevick, G. R. Jelbert, M. Aykol, and C. Wolverton, Acta Mater. **159**, 102 (2018).
[11]Y. T. Sun, H. Y. Bai, M. Z. Li, and W. H. Wang, J. Phys. Chem. Lett. **8**, 3434 (2017).
[12]A. Majid, S. B. Ahsan, N. ul, and H. Tariq, Appl. Soft Comput. **28**, 569 (2015).
[13]J. Xiong, S.-Q. Shi, and T.-Y. Zhang, Mater. Des. **187**, 108378 (2020).
[14]D. G. Mastropietro and J. A. Moya, Comput. Mater. Sci. **188**, 110230 (2021).
[15]M. Samavatian, R. Gholamipour, and V. Samavatian, Comput. Mater. Sci. **186**, 110025 (2021).
[16]Y. Zhang, H. Tan, H. Z. Kong, B. Yao, and Y. Li, J. Mater. Res. **18**, 664 (2003).
[17]B. Deng and Y. Zhang, Chem. Phys. **538**, 110898 (2020).
[18]B. Ren, Z. Long, and R. Deng, Comput. Mater. Sci. **189**, 110259 (2021).
[19]Y.-Y. Chen, Y.-H. Lin, C.-C. Kung, M.-H. Chung, and I.-H. Yen, Sensors **19**, 2047 (2019).
[20]J. Xiong, S.-Q. Shi, and T.-Y. Zhang, Comput. Mater. Sci. **192**, 110362 (2021).
[21]Z. P. Lu, Y. Li, and S. C. Ng, J. Non-Cryst. Solids **270**, 103 (2000).
[22]K. Mondal and B. S. Murty, J. Non-Cryst. Solids **351**, 1366 (2005).
[23]B. Dong, S. Zhou, D. Li, C. Lu, F. Guo, X. Ni, and Z. Lu, Prog. Natl. Sci. Mater. Int. **21**, 164 (2011).
[24]G. J. Fan, H. Choo, and P. K. Liaw, J. Non-Cryst. Solids **353**, 102 (2007).
[25]X. Xiao, F. Shoushi, W. Guoming, H. Qin, and D. Yuanda, J. Alloys Compd. **376**, 145 (2004).
[26]S. Guo and C. T. Liu, Intermetallics **18**, 2065 (2010).
[27]M. K. Tripathi, S. Ganguly, P. Dey, and P. P. Chattopadhyay, Comput. Mater. Sci. **118**, 56 (2016).
[28]Z. Long, G. Xie, H. Wei, X. Su, J. Peng, P. Zhang, and A. Inoue, Mater. Sci. Eng., A **509**, 23 (2009).
[29]D. R. Uhlmann, J. Non-Cryst. Solids **7**, 337 (1972).
[30]Q. Dong, Y. J. Pan, J. Tan, X. M. Qin, C. J. Li, P. Gao, Z. X. Feng, M. Calin, and J. Eckert, J. Alloys Compd. **785**, 422 (2019).
[31]M. Mohammadi Rahvard, M. Tamizifar, and S. M. A. Boutorabi, J. Non-Cryst. Solids **481**, 74 (2018).
[32]M. Mohammadi Rahvard, M. Tamizifar, and S. M. A. Boutorabi, J. Therm. Anal. Calorim. **134**, 903 (2018).