

## ARTICLE OPEN

Cluster-formula-embedded machine learning for design of multicomponent  $\beta$ -Ti alloys with low Young's modulusFei Yang<sup>1</sup>, Zhen Li<sup>2</sup>✉, Qing Wang<sup>3</sup>✉, Beibei Jiang<sup>3</sup>, Biaojie Yan<sup>4</sup>, Pengcheng Zhang<sup>4</sup>, Wei Xu<sup>5</sup>, Chuang Dong<sup>1</sup> and Peter K. Liaw<sup>6</sup>

The present work formulated a materials design approach, a cluster-formula-embedded machine learning (ML) model, to search for body-centered-cubic (BCC)  $\beta$ -Ti alloys with low Young's modulus ( $E$ ) in the Ti–Mo–Nb–Zr–Sn–Ta system. The characteristic parameters, including the Mo equivalence and the cluster-formula approach, are implemented into the ML to ensure the accuracy of prediction, in which the former parameter represents the BCC- $\beta$  structural stability, and the latter reflects the interactions among elements expressed with a composition formula. Both auxiliary gradient-boosting regression tree and genetic algorithm methods were adopted to deal with the optimization problem in the ML model. This cluster-formula-embedded ML can not only predict alloy property in the forward design, but also design and optimize alloy compositions with desired properties in multicomponent systems efficiently and accurately. By setting different objective functions, several new  $\beta$ -Ti alloys with either the lowest  $E$  ( $E = 48$  GPa) or a specific  $E$  ( $E = 55$  and  $60$  GPa) were predicted by ML and then validated by a series of experiments, including the microstructural characterization and mechanical measurements. It could be found that the experimentally obtained  $E$  of predicted alloys by ML could reach the desired objective  $E$ , which indicates that the cluster-formula-embedded ML model can make the prediction and optimization of composition and property more accurate, effective, and controllable.

npj Computational Materials (2020)6:101; <https://doi.org/10.1038/s41524-020-00372-w>

## INTRODUCTION

$\beta$ -Ti alloys with a body-centered-cubic (BCC) structure have attracted more attention due to their prominent properties (high strength, low Young's modulus ( $E$ ), and good corrosion resistance), showing a great potential for biomedical applications<sup>1,2</sup>. However, the low Young's modulus and BCC- $\beta$  structural stability are a pair of contradiction, which is difficult to be solved in simple alloy systems. Low- $E$   $\beta$ -Ti alloys with  $E = 40$ – $65$  GPa are generally formed in multicomponent systems, such as Ti–35Nb–5Ta–7Zr ( $E = 55$  GPa) and Ti–24Nb–4Zr–7.9Sn ( $E = 42$  GPa, the number before each element represents the weight percent, wt. %) <sup>2–7</sup>. The BCC- $\beta$  structural stabilities of these alloys are always in the vicinity of the lower limit of  $\beta$  stabilization, which was often characterized by a Mo-equivalent parameter ( $Mo_{eq}$ )<sup>8,9</sup>. Besides  $Mo_{eq}$ , the  $d$ -electron theory method<sup>10,11</sup> and valence electron concentration<sup>12–14</sup> are frequently used to design multicomponent low- $E$  alloys. A cluster-formula approach is also applied to design alloy compositions in multicomponent systems, which considers the correlations among alloying solute elements and the base element in light of the chemical short-range orders (CSROs) of solid solutions and then gives chemical compositions intuitively<sup>15–17</sup>. Thus, a series of low- $E$   $\beta$ -Ti alloys have been achieved by the cluster formula of  $[(Mo,Sn) - (Ti,Zr)_{14}](Nb,Ta)_{1-3}$ , in which the lowest  $E = 48$  GPa could reach at the  $[(Mo_{0.5}Sn_{0.5}) - (Ti_{13}Zr_1)]Nb_1$  alloy ( $Ti_{81.25}Zr_{6.25}Nb_{6.25}Mo_{3.125}Sn_{3.125}$  in atomic percent at. %) <sup>18</sup>. Moreover, these alloy compositions were restrained further by a modified  $Mo_{eq}$  parameter, and it was found that the  $Mo_{eq}$  values of these low- $E$  alloys are very close to the lower limit of  $\beta$  stabilization (11.8 wt.% Mo)<sup>19</sup>. A higher  $Mo_{eq}$  corresponds to a relatively high  $E$ , and a relatively lower  $Mo_{eq}$  could not render alloys with a single

BCC- $\beta$  structure, which induces a high  $E$  due to the precipitation of second phases, such as  $\omega$  and  $\alpha''$ <sup>20</sup>.

Although the experimental trial-and-error approaches based on several physical-metallurgy-guided rules could design and obtain new materials with targeted properties, it could not consider issues comprehensively and efficiently<sup>21,22</sup>. Recently, artificial intelligence methods, especially the machine learning (ML), have attracted more attention since these methods could predict alloy properties and design new high-performance materials efficiently and accurately with the assistance of physical-metallurgy rules<sup>23–28</sup>. For instance, the ML model could establish a relationship between the input and output through training data, in which the predicted capability strongly depends on the size and quality of the database as well as the range and distribution of the input parameters. Since these databases generally contain small data, the predictions that rely on ML alone may not be accurate enough. Furthermore, complex interactions among alloying elements will also affect the predicted accuracy. Therefore, the ML method should be combined with some characteristic parameters in a given system, which could result in more effective predictions. Xue et al. carried out the ML surrogate model, together with characteristic parameters (such as itinerant electron concentration  $e/a$ , modulus mismatch  $\eta$ , and work function  $w$ , etc.), to search for high-entropy alloys with high microhardness in Al–Co–Cr–Cu–Fe–Ni system, in which the predicted results are well consistent with the experiments since these selected characteristic parameters are closely related to the desired property<sup>29</sup>. Similarly, the combination of structural and compositional features (cohesive energy, atomic radius, and electronegativity) with the ML could well predict the elastic properties of

<sup>1</sup>Key Laboratory of Materials Modification by Laser, Ion and Electron Beams (Ministry of Education), School of Materials Science and Engineering, Dalian University of Technology, Dalian, 116024 Liaoning, China. <sup>2</sup>School of Mechanical Engineering, Dalian University of Technology, Dalian, 116024 Liaoning, China. <sup>3</sup>Analysis and Test Center, Guangdong University of Technology, Guangzhou, 510006 Guangdong, China. <sup>4</sup>Institute of Materials, China Academy of Engineering Physics, Jiangyou, 621908 Sichuan, China. <sup>5</sup>State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang, 110819 Liaoning, China. <sup>6</sup>Department of Materials Science and Engineering, The University of Tennessee, Knoxville, 37996 TN, USA. ✉email: lizhen@dlut.edu.cn; wangq@dlut.edu.cn

Al–Co–Cr–Fe–Ni high-entropy alloys<sup>30</sup>. Xu et al. embedded the physical-metallurgy parameters, the volume fraction, and driving force of second-phase precipitates that represent the microstructural features, into the ML model to design advanced ultrahigh-strength stainless steels successfully<sup>31</sup>. The optimal prototype alloy predicted by this model was well demonstrated by experiments, in which an excellent agreement could be obtained for the predicted optimal parameter settings and the final mechanical property (microhardness). In fact, the introduction of any characteristic parameters in the ML process aims mainly at constructing a perfect target-oriented loop of the input (alloy compositions) and output (properties) for an accurate and efficient prediction and design.

It is noted that any composition formula has not been considered in the ML method until now. Therefore, the present work will embed a cluster-formula approach into the ML model in multicomponent low- $E$   $\beta$ -Ti alloys because this approach is closely related to alloy compositions, besides the  $Mo_{eq}$  parameter. In the forward design from the input (alloy compositions,  $c_i$ ) to output (Young's modulus,  $E$ ), the auxiliary gradient-boosting regression tree (XGBoost) method<sup>32</sup>, random forest (RF) method<sup>33</sup>, and support vector regression (SVR) method<sup>34</sup> will be applied to establish the relationship between the composition and Young's modulus based on the existing experimental results in Ti–Mo–Nb–Zr–Sn–Ta systems, including binary, ternary, quaternary, and quinary low- $E$   $\beta$ -Ti alloys. The  $Mo_{eq}$  parameter will be embedded into this process to constrain the BCC- $\beta$  structural stability. Then, in the reverse design of an alloy composition according to a specific  $E$  value, the cluster-formula approach will also be embedded into the ML model to simplify the selection of alloy compositions since a given  $E$  value would correspond to many compositions. Moreover, the genetic algorithm (GA) will be adopted to deal with the nonlinear alloy optimization problem. Finally, the predicted alloys with the cluster-formula-embedded ML model will be validated through a series of experiments, including microstructural characterizations and mechanical measurements.

## RESULTS

### Characteristic parameters in low- $E$ $\beta$ -Ti alloys

Among all the parameters to characterize the structural stability of the metastable  $\beta$ -Ti, the Mo equivalence ( $Mo_{eq}$ ) is the most popular parameter in multicomponent low- $E$   $\beta$ -Ti alloys. It is expressed with the equation of

$$Mo_{eq} = 1.0\text{Mo} + 1.25\text{V} + 0.59\text{W} + 0.33\text{Nb} + 0.25\text{Ta} + 1.93\text{Fe} \\ + 1.84\text{Cr} + 1.50\text{Cu} + 2.46\text{Ni} + 2.67\text{Co} + 2.26\text{Mn} \\ + 0.30\text{Sn} + 0.31\text{Zr} + 3.01\text{Si} - 1.47\text{Al} (\text{in wt.}\%) \quad (1)$$

in which the coefficient before each alloying element was obtained by the slope ratio of the  $[\beta/(\alpha + \beta)]$  phase boundary of the binary Ti–M phase diagram to that of the Ti–Mo, representing the contribution of each element to the  $\beta$  structural stability<sup>9,19</sup>. The critical lower limit of  $\beta$  stabilization is determined as  $(Mo_{eq})_C = 11.8$  wt.%, indicating that any alloy with a larger  $Mo_{eq}$  value above 11.8 wt.% would exhibit a single BCC  $\beta$ -Ti structure<sup>9,19</sup>. It is known that in the vicinity of the lower limit of  $\beta$  stabilization, some second phases of  $\omega$ ,  $\alpha''$ , and  $\alpha'$  could be precipitated inevitably since they are sensitive to the preparation processing and heat treatments, resulting in an increase of  $E$ <sup>20</sup>. Experimentally, only when the  $Mo_{eq}$  value is larger than 13.0 wt.%, these second phases could be avoided and the  $E$  could reach the minimum<sup>9,35</sup>. It is noted that the  $Mo_{eq}$  parameter ignores the interactions among elements, which is an important factor because the excessive addition of elements with strong

interactions could result in the formation of intermetallic compounds in solid-solution alloys<sup>9,20</sup>. Furthermore, the  $Mo_{eq}$  describes the  $\beta$  structural stability of Ti alloys alone, and could not correlate directly with the properties, such as the Young's modulus.

Actually, the interactions among alloying elements with the base element could be incarnated into the CSROs of solid solutions, which is the significant microstructural feature<sup>15–17</sup>. Based on CSROs, we proposed a 'cluster-plus-glue-atom' structural model to describe the local atomic distribution of alloying solute elements. In this model, the cluster is the nearest-neighbor polyhedron centered by a solute atom that has strong interaction (characterized by a large negative enthalpy of mixing ( $\Delta H$ )<sup>36</sup>) with the base solvent atoms to represent the strongest CSRO. Some other solute atoms having weak interactions with the base (a positive  $\Delta H$ ) are certainly required to fill the space among clusters for the balance of atomic-packing density, named as glue atoms<sup>37</sup>. Thus, an uniform composition formula [cluster](glue atom) <sub>$m$</sub>  ( $m$  being the glue-atom number) can be abstracted from this model, which is called the cluster-formula approach<sup>18,19</sup>. Specifically, in the Ti–Mo–Nb–Zr–Sn–Ta multicomponent system, Mo and Sn occupy preferentially the cluster center due to the relatively strong interactions with Ti ( $\Delta H_{\text{Ti–Mo}} = -4$  kJ/mol,  $\Delta H_{\text{Ti–Sn}} = -21$  kJ/mol), while Nb and Ta tend to occupy the glue sites owing to their weak interactions with Ti ( $\Delta H_{\text{Ti–Nb}} = +2$  kJ/mol,  $\Delta H_{\text{Ti–Ta}} = +1$  kJ/mol). For the Zr element, it can enter into any site to replace the Ti since they are in the same group ( $\Delta H_{\text{Ti–Zr}} = 0$  kJ/mol). Thus, an ideal cluster formula could be expressed as  $[(\text{Mo}_u\text{Sn}_v) - (\text{Ti}_w\text{Zr}_{z_1})] (\text{Nb}_x\text{Ta}_y\text{Zr}_{z_2}) (x + y + z_3 = m = 1, 2, 3)$

$$[(\text{Mo}_u\text{Sn}_v) - (\text{Ti}_w\text{Zr}_{z_1})] (\text{Nb}_x\text{Ta}_y\text{Zr}_{z_2}) (x + y + z_3 = m = 1, 2, 3) \quad (2)$$

in which the total amount ( $x + y + z_3$ ) of Nb, Ta, and Zr (i.e., the glue-atom number,  $m$ ) varies from  $m = 1$  to  $m = 3$  according to the Mo content ( $u$ ) since Mo is a strong  $\beta$ -stabilized element; the amount of Mo and Sn in the cluster center is  $u + v = 1$  ( $0 \leq v < 1$ ); the amount of the base Ti and partial Zr in the cluster shell is defined as  $w + z_1 = 14$  ( $0 \leq z_1 \leq 1$ ); and the total amount of Zr is the sum of that in the cluster shell and glue site, being  $z = z_1 + z_3$ . It is noted that the glue atoms contain the Zr element occasionally for the consistency with the existing experimental results of Zr substitution for Nb.

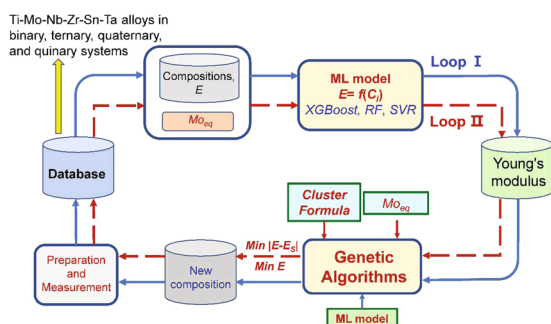
$$[(\text{Zr}_{z_2}\text{Sn}_v) - (\text{Ti}_w\text{Zr}_{z_1})] (\text{Nb}_x\text{Ta}_y) (x + y = m = 3, 4, 5) \quad (3)$$

in which the glue-atom number ( $x + y$ ) of Nb and Ta should be increased because Nb and Ta are weak  $\beta$ -stabilizers when the Mo element is not added; the amount of Sn and partial Zr in the cluster center is  $z_2 + v = 1$  ( $0 \leq v < 1$ ); the amount of the base Ti and partial Zr in the cluster shell is  $w + z_1 = 14$  ( $0 \leq z_1 \leq 1$ ); the total Zr content is  $z = z_1 + z_2$ .

These concrete composition formulas will be then embedded into the ML to constrain complex compositions during the reverse design process in the present work.

### Machine-learning model

A ML model will be trained based on a dataset containing Young's modulus ( $E$ ) and alloy composition ( $c_i$ ) for each element to build the relationship of  $E = f(c_i)$  in the present work, as seen in Fig. 1. Thus, the  $E$  value of any alloy could be predicted when the composition is known in the forward design (Loop I in Fig. 1). On the other side, the prediction of an alloy composition could also be realized if the  $E$  value is given in the reverse design (Loop II in Fig. 1). In order to make the prediction more accurate, some characteristic parameters representing the low- $E$  and  $\beta$  structural



**Fig. 1 Schematic diagram of the design of multicomponent  $\beta$ -Ti alloys with low- $E$  based by machine learning.** XGBoost, RF, and SVR models are trained by the database constituted by Ti-Mo-Nb-Zr-Sn-Ta alloys, containing alloy composition,  $E$  and  $Mo_{eq}$  in the forward design (Loop I) from the input (composition) to the output ( $E$ ). While in the reverse design (Loop II), the genetic algorithm is used to solve the optimization problems ( $Min E$  and  $Min |E - E_s|$ ), during which the cluster-formula approach is embedded to further constraint alloy compositions for a specific  $E$ , combined with  $Mo_{eq}$ .

stability should be set into the ML process to constrain the relationship of  $E = f(c_i)$ , since it is a function of multiple variables for a single target. For the dataset, it is constituted of 82 existing low- $E$   $\beta$ -Ti alloy samples<sup>3,5,7,9-11,18,19,38-59</sup>, including binary, ternary, quaternary, and quinary alloys in Ti-Mo-Nb-Zr-Sn-Ta systems, which are listed in Supplementary Table 1. Sample data were screened strictly with the  $Mo_{eq}$  parameter to avoid the effects of second-phase precipitation and processing conditions on  $E$ . That is to say, alloys containing only a minor amount of  $\alpha''$  and  $\omega$ , which was checked scarcely, were permitted in the dataset since the obvious phase precipitation could enhance the  $E$ . Besides, all these alloys were prepared by rapid-quenching processing.

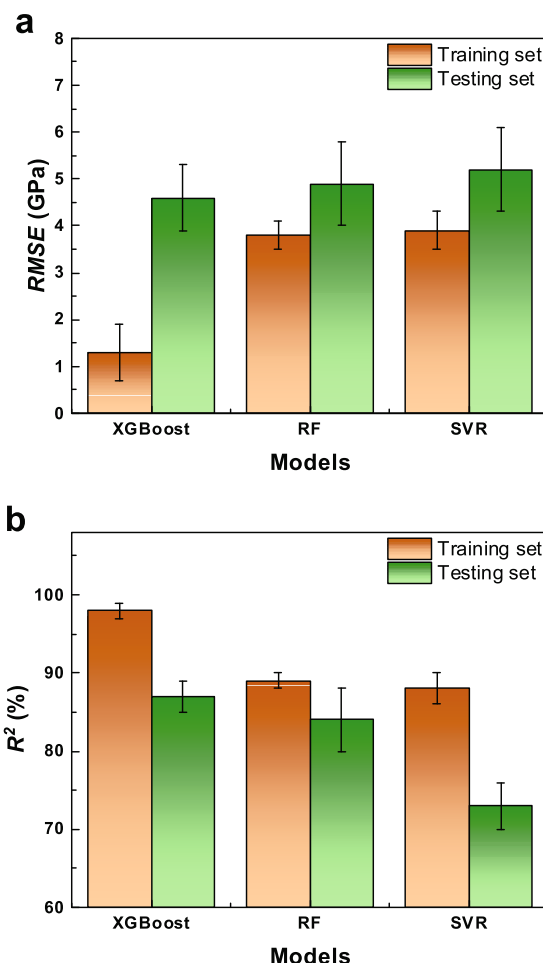
In the forward design of ML (Loop I in Fig. 1), we employed three representative ML models, XGBoost implemented in the XGBoost machine-learning libraries<sup>32</sup>, RF, and SVR with a radial basis function kernel implemented in the scikit-learn package<sup>34</sup>. During the ML process, the characteristic  $Mo_{eq}$  parameter was also put into the dataset to constrain alloy compositions further because it represents the  $\beta$  structural stabilities of Ti alloys. Then, this dataset was split into a training set and a testing set by using a split ratio of 9:1 for the generalization ability of the models, i.e., the training set was taken from 90% of the dataset. We used both the root mean-squares error ( $RMSE$ ) and the coefficient of determination ( $R^2$ ) as the criterion for the prediction accuracy. The  $RMSE$  and  $R^2$  are expressed with the equations of

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2} \quad (4)$$

and

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f(x_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (5)$$

respectively, where  $n$  is the number of samples,  $\bar{y}$  is the mean Young's modulus of samples,  $y_i$  and  $f(x_i)$  are the experimental Young's modulus and the predicted value by ML, respectively. Since the dataset capacity of 82 samples is somewhat limited in the present work, it is necessary to employ the multiple hold-out method to calculate the  $RMSE$  and  $R^2$  of ML models for the guarantee of accuracy<sup>31</sup>. This procedure was repeated for 500 times in these three models by partitioning the training set and testing set randomly with a ratio of 9/1, and then the  $RMSE$  and  $R^2$  values were counted, as shown in Fig. 2. The mean values of  $RMSE$  of the training set are  $1.3 \pm 0.6$ ,  $3.8 \pm 0.3$ , and  $3.9 \pm 0.4$  GPa for



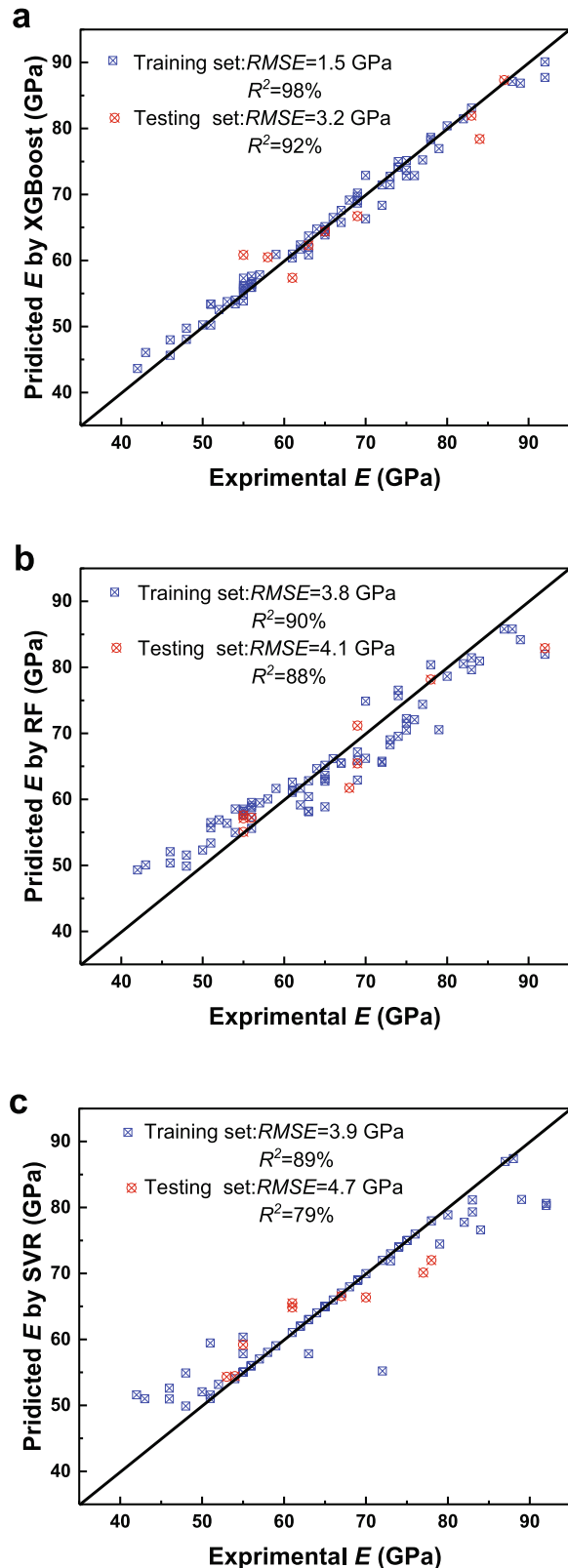
**Fig. 2 Machine-learning models and their statistical results on both training set and testing set. a  $RMSE$ . b  $R^2$ .** The error bars are the standard deviation from the predictions of 500 regression models.

XGBoost, RF, and SVR models, respectively; and these values of the testing set are  $4.6 \pm 0.7$ ,  $4.9 \pm 0.9$ , and  $5.2 \pm 0.9$  GPa for XGBoost, RF, and SVR, respectively. Besides, the mean values of  $R^2$  of the training set are  $98 \pm 1$ ,  $89 \pm 1$ , and  $88 \pm 2\%$  for XGBoost, RF, and SVR, respectively; and these values of the testing set are  $87 \pm 2$ ,  $84 \pm 4$ , and  $73 \pm 3\%$ , respectively. Thereof, both the predicted accuracy and generalized ability of the XGBoost model are better than the RF and SVR models for the prediction of alloy property.

Figure 3 gives the optimal results for both the training and testing sets trained by the XGBoost, RF, and SVR models, respectively, in which the blue and red points represent the training set and the testing set. The  $RMSE$  values of the training and testing sets for XGBoost are 1.5 and 3.2 GPa, respectively, while the  $RMSE$  values of the training and testing sets for RF and SVR are 3.8 and 4.1 GPa, 3.9 and 4.7 GPa, respectively. Meanwhile, the  $R^2$  values of the training and testing sets for XGBoost are 98 and 92%, while the corresponding  $R^2$  values are 90 and 88% for RF, and 89 and 79% for SVR, respectively. It could be found that most of the predicted values by the XGBoost model in both the training and testing sets are more consistent with the experimental results than the RF and SVR predictions, indicating that the XGBoost prediction is more accurate than the RF and SVR.

Once the relationship between the alloy composition and Young's modulus was established by the XGBoost model in Loop I, the Young's modulus of any given composition alloy could be well predicted. However, in the reverse design of ML, a given  $E$  value

might correspond to dozens or hundreds of multicomponent compositions, which would certainly lead to a complexity in the prediction of low- $E$   $\beta$ -Ti alloys. In order to solve this problem, we



**Fig. 3** The experimental  $E$  values vs. the predicted values by the optimal model. **a** XGBoost. **b** RF. **c** SVR. The blue points and red points represent the training set and the testing set, respectively.

implemented the cluster-formula approach into the reverse design (Loop II in Fig. 1) to constrain the correlations among alloying elements with the aid of their interactions with the base Ti, besides the  $Mo_{eq}$  parameter. Thus, a given  $E$  value will correspond to very few compositions alone in a given system, which could reduce the difficulty of composition exploration drastically. Thereof, this problem can be converted into an optimization problem by virtue of the cluster-composition formula:

$$\begin{aligned} &\text{Minimize} && E(x, y, z, u, v, w) \\ &\text{subject to} && CF(x, y, z, u, v, w) \\ &&& 11.0 \leq Mo_{eq}(x, y, z, u, v, w) \leq 14.0 \end{aligned} \quad (6)$$

Or

$$\begin{aligned} &\text{Minimize} && |E(x, y, z, u, v, w) - E_s| \\ &\text{subject to} && CF(x, y, z, u, v, w) \\ &&& 11.0 \leq Mo_{eq}(x, y, z, u, v, w) \leq 14.0 \end{aligned} \quad (7)$$

where  $x, y, z, u, v$ , and  $w$  represent the atom numbers (or the amounts) of Nb, Ta, Zr, Mo, Sn, and Ti in a given cluster-composition formula, respectively;  $E(x, y, z, u, v, w)$  is the predicted Young's modulus trained by the XGBoost model;  $CF(x, y, z, u, v, w)$  is the specific cluster-composition formula given in above-mentioned Eqs. (2) and (3). Here, we took the strengthen elitist selection of the GA<sup>60–63</sup> to solve this optimization problem. For the optimization problem given in Eq. (6), we aim at exploring of  $\beta$ -Ti alloys with the lowest Young's modulus ( $Min E$ ). While for a specific  $E_s$  value, it is necessary to minimize the difference between  $E(x, y, z, u, v, w)$  and  $E_s$ , i.e.,  $Min |E - E_s|$ , as described in the optimization problem given in Eq. (7).

Thus, an integrated ML process would be established, as presented in Fig. 1. According to it, the predicted new composition alloys will be verified by a series of experiments, then they will be put into the original database for a much more accurate prediction.

#### Experimental verification

Firstly, the XGBoost model is used to predict the Young's modulus  $E$  of any given alloy in Ti–Mo–Nb–Zr–Sn–Ta system in the forward design of ML. We selected two alloys randomly only with the  $Mo_{eq}$  constraint,  $Ti_{84.5}Mo_{2.0}Nb_{10.0}Sn_{1.0}Ta_{2.5}$  (I-1, at. %) and  $Ti_{82.5}Mo_{3.0}Nb_{10.5}Sn_{1.0}Ta_{3.0}$  (I-2, at. %). The  $E$  values of these two alloys predicted by ML is  $E = 67$  GPa for I-1 and  $E = 65$  GPa for I-2 alloy, respectively (as seen in Table 1). Then, a series of experimental measurements were carried out to verify this prediction.

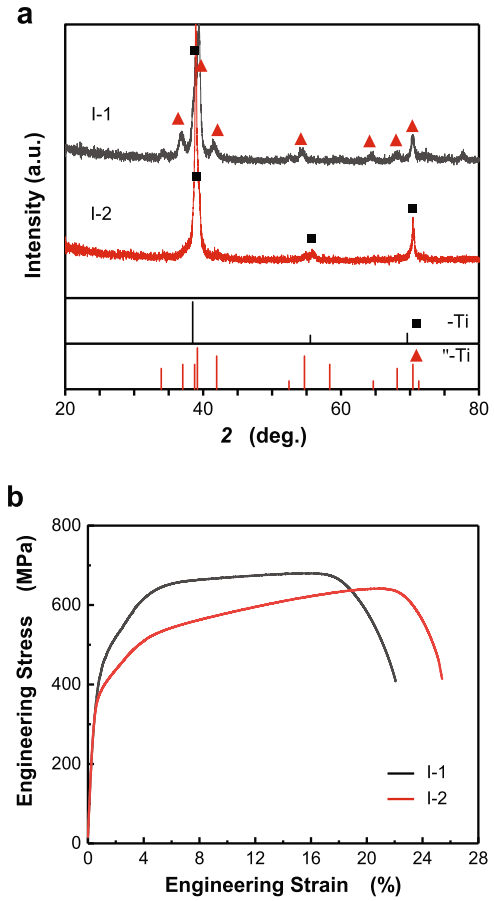
The X-ray diffractometer (XRD) results in Fig. 4a show that the I-2 alloy with  $Mo_{eq} = 13.30$  wt.% exhibits a single BCC- $\beta$  structure, while the diffraction peaks of the  $\alpha''$  martensite with an orthorhombic structure appear in I-1 alloy, besides the BCC- $\beta$ . It is resulted from that the  $Mo_{eq} = 11.28$  wt.% of I-1 is relatively lower than the critical lower limit ( $(Mo_{eq})_C = 11.8$  wt.%) for the  $\beta$  stabilization. The tensile tests of these alloys were then measured, and the engineering tensile stress-strain curves were given in Fig. 4b, from which the Young's modulus could be calculated. The experimental  $E$  values of I-1 and I-2 alloys are  $E = 76 \pm 3$  and  $E = 73 \pm 2$  GPa, respectively, which is somewhat different from the corresponding predicted values (67 and 65 GPa). The difference between the predicted and experimental  $E$  values might be ascribed to that alloy samples in this quinary system were not included in the dataset due to no existing results, which would give rise to an uncertainty and then reduce the prediction accuracy. In addition, the higher measured  $E$  value of I-1 alloy is also resulted from the obvious precipitation of  $\alpha''$  martensite that could increase the  $E$  slightly. Thereof, the predication of  $E$  is not accurate enough when the  $Mo_{eq}$  is considered in ML alone. There



**Table 1.** The alloys predicted by machine learning and their predicted and experimental Young's modulus ( $E$ ).

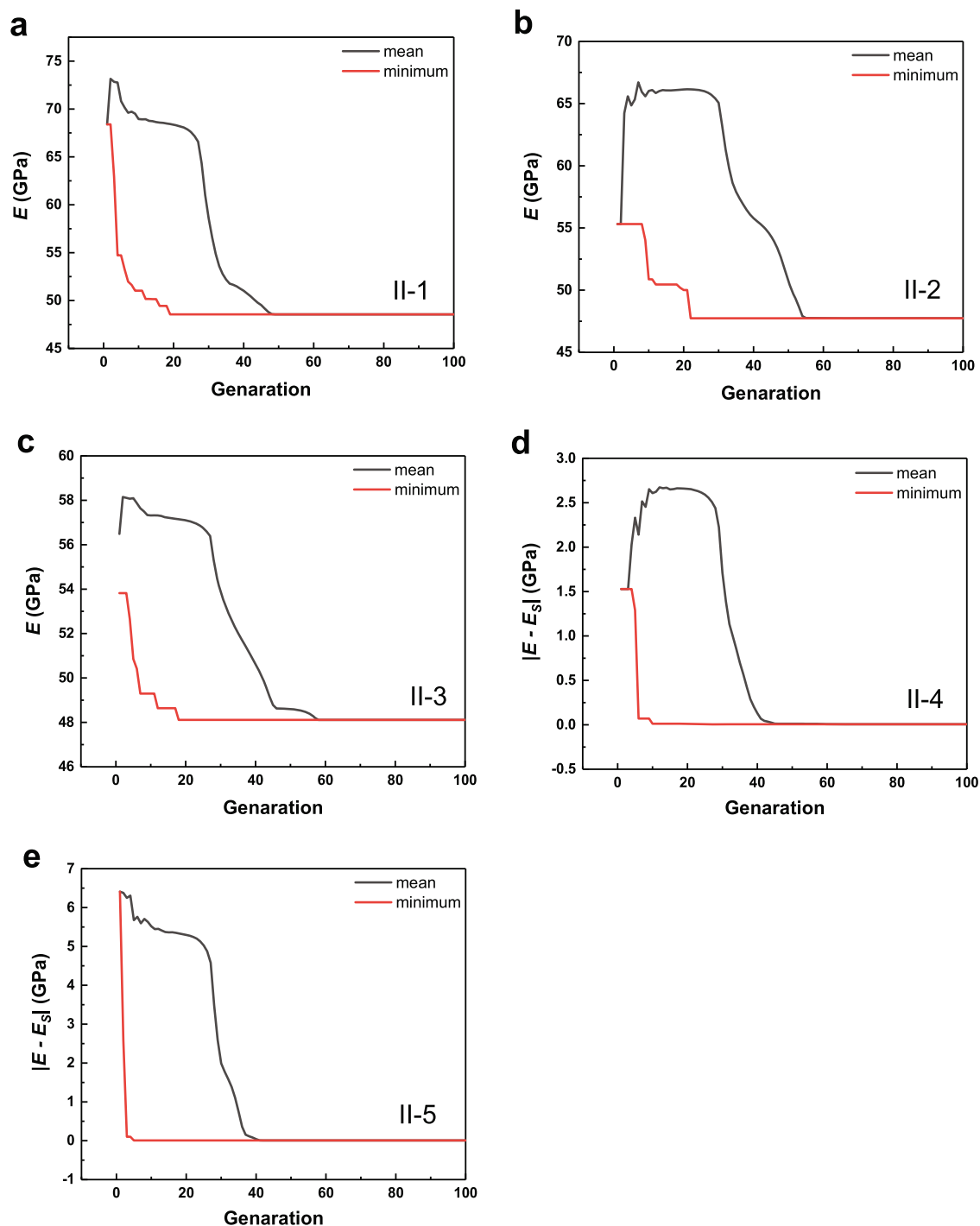
Alloy no.	Cluster formulas	Compositions (at. %)	(wt. %)	$Mo_{eq}$ (wt. %)	Experimental $E$ (GPa)	Predicted $E$ (GPa)
I-1	-	$Ti_{84.5}Mo_2Nb_{10}Sn_1Ta_{2.5}$	Ti-3.3Mo-16.2Nb-2.1Sn-7.9Ta	11.28	$76 \pm 3$	67
I-2	-	$Ti_{82.5}Mo_3Nb_{10.5}Sn_1Ta_3$	Ti-4.9Mo-16.6Nb-2.5Sn-9.2Ta	13.30	$73 \pm 2$	65
II-1	$[(Mo_{0.5}Sn_{0.5}) - (Ti_{1.3}Zr_{1.1})] (Nb_{0.5}Ta_{0.5})$	$Ti_{81.3}Mo_{3.1}Nb_{3.1}Zr_{6.3}Sn_{3.1}Ta_{3.1}$	Ti-5Mo-4.9Nb-9.5Zr-6.2Sn-9.4Ta	13.78	$49 \pm 1$	49
II-2	$[(Mo_{0.3}Sn_{0.7}) - (Ti_{1.3.5}Zr_{0.5})] (Nb_{1.5}Ta_{0.5})$	$Ti_{79.4}Mo_{1.8}Nb_{8.8}Zr_{2.9}Sn_{4.1}Ta_{2.9}$	Ti-2.8Mo-13.5Nb-4.4Zr-8Sn-8.8Ta	13.20	$46 \pm 1$	48
II-3	$[(Sn_{0.5}Zr_{0.5}) - (Ti_{1.4})] (Nb_{2.5}Ta_{0.5})$	$Ti_{77.8}Nb_{13.9}Zr_{2.8}Sn_{2.8}Ta_{2.8}$	Ti-21.2Nb-4.2Zr-5.4Sn-8.2Ta	11.95	$47 \pm 2$	48
II-4	$[(Mo_{0.5}Sn_{0.5}) - (Ti_{1.3}Zr_{1.1})] (Nb_{0.8}Zr_{0.2})$	$Ti_{81.3}Mo_{3.1}Nb_5Zr_{7.5}Sn_{3.1}$	Ti-5.3Mo-8.1Nb-12Zr-6.5Sn	13.60	$58 \pm 2$	55
II-5	$[(Mo_{0.5}Sn_{0.5}) - (Ti_{1.3}Zr_{1.1})] (Nb_{0.5}Zr_{0.5})$	$Ti_{81.3}Mo_{3.1}Nb_{3.1}Zr_{9.4}Sn_{3.1}$	Ti-5.3Mo-5.1Nb-15Zr-6.5Sn	13.53	$56 \pm 2$	60

Two alloy compositions (I-1 and I-2) with both atomic percent and weight percent expressions are selected with the  $Mo_{eq}$  constraint in the forward design. Five alloy compositions (II-1–II-5) are predicted with the cluster-formula constraint in the reverse design, which are expressed with both cluster formula and atomic percent. The Mo equivalence is expressed with  $Mo_{eq} = 1.0Mo + 0.33Nb + 0.30Sn + 0.31Zr + 0.25Ta$  (wt. %).

**Fig. 4** Experimental results of  $Ti_{84.5}Mo_2Nb_{10.0}Sn_{1.0}Ta_{2.5}$  (I-1, at. %) and  $Ti_{82.5}Mo_3Nb_{10.5}Sn_{1.0}Ta_{3.0}$  (I-2, at. %) alloys. **a** XRD patterns. **b** Engineering tensile stress-strain curves. The I-2 alloy exhibits a single BCC- $\beta$  structure, while the I-1 alloy consists of the BCC- $\beta$  and orthorhombic  $\alpha'$  martensite.

must exist some inevitable uncertainties in the ML process, which needs to consider more other parameters for a further constraint of alloy compositions, such as the factor to reflect complex interactions among elements in multicomponent systems. Thereof, we add the cluster formulas as a constraint in the reverse design (Loop II) to explore alloy composition according to a given  $E$  value.

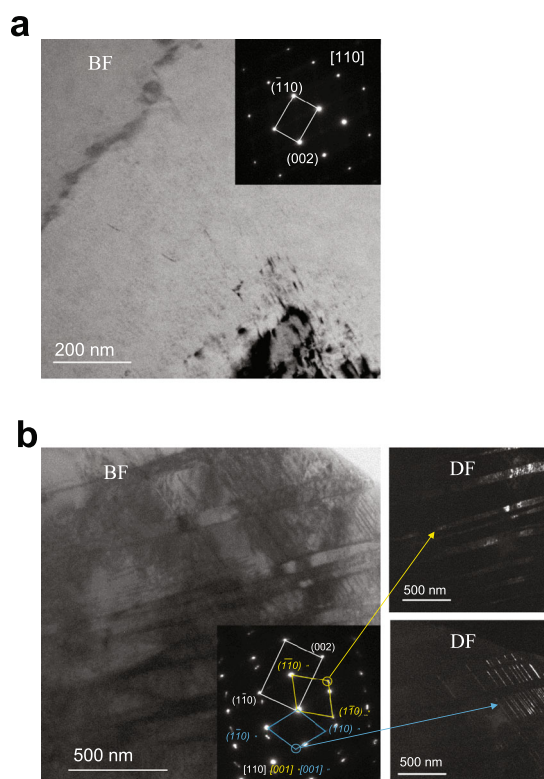
During the process of the GA optimization in the reverse design (Loop II), we still use the XGBoost model in the forward design and take the  $Mo_{eq}$  and the cluster formulas as constraints simultaneously. A wide composition range of each element was first set, in which a population containing 3000000 individuals (i.e., alloy compositions) was generated randomly. Then, the composition alloys satisfying both the cluster formula and the  $Mo_{eq}$  parameters were obtained with the XGBoost model, and the fitness of each alloy composition was calculated according to the objective function of  $Min E$  for the optimization problem given in Eq. (6) or  $Min |E - E_s|$  for the optimization problem given in Eq. (7). Thus, the next generation of population was produced by the genetic operations (including selection, crossover, and mutation) in light of the survival of the fittest principle. Finally, the GA would terminate after 100 iterations, and the optimal alloy composition could be achieved. Figure 5 gives the evolution histories of the GA with the objective functions of  $Min E$  (Fig. 5a–c) and  $Min |E - E_s|$  (Fig. 5d, e), in which the red line represents the objective function while the black line represents the mean value of  $E$  or  $|E - E_s|$  for each generation. It was found that these two values will be converged after about 50 iterations, indicating a high efficiency of



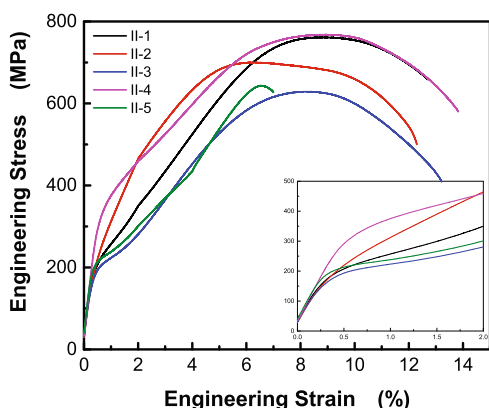
**Fig. 5** The evolution histories of the GA with the objective functions of  $\text{Min } E$  and  $\text{Min } |E - E_s|$ . **a**  $[(\text{Mo}_{0.5}\text{Sn}_{0.5}) - (\text{Ti}_{13}\text{Zr}_1)](\text{Nb}_{0.5}\text{Ta}_{0.5})$  (II-1). **b**  $[(\text{Mo}_{0.3}\text{Sn}_{0.7}) - (\text{Ti}_{13.5}\text{Zr}_{0.5})](\text{Nb}_{1.5}\text{Ta}_{0.5})$  (II-2). **c**  $[(\text{Sn}_{0.5}\text{Zr}_{0.5}) - \text{Ti}_{14}](\text{Nb}_{2.5}\text{Ta}_{0.5})$  (II-3). **d**  $[(\text{Mo}_{0.5}\text{Sn}_{0.5}) - (\text{Ti}_{13}\text{Zr}_1)](\text{Nb}_{0.8}\text{Zr}_{0.2})$  (II-4). **e**  $[(\text{Mo}_{0.5}\text{Sn}_{0.5}) - (\text{Ti}_{13}\text{Zr}_1)](\text{Nb}_{0.5}\text{Zr}_{0.5})$  (II-5). The red line represents the value of the objective function while the black line represents the mean value of the  $E$  or  $|E - E_s|$  for each generation. Five alloy compositions presented in (a–e) are obtained finally.

this optimization. So, three new alloys with the cluster formulas of  $[(\text{Mo}_{0.5}\text{Sn}_{0.5}) - (\text{Ti}_{13}\text{Zr}_1)](\text{Nb}_{0.5}\text{Ta}_{0.5})$  (II-1),  $[(\text{Mo}_{0.3}\text{Sn}_{0.7}) - (\text{Ti}_{13.5}\text{Zr}_{0.5})](\text{Nb}_{1.5}\text{Ta}_{0.5})$  (II-2), and  $[(\text{Sn}_{0.5}\text{Zr}_{0.5}) - \text{Ti}_{14}](\text{Nb}_{2.5}\text{Ta}_{0.5})$  (II-3) are obtained by setting the objective function of  $\text{Min } E$ , being  $E = 49$ ,  $E = 48$ , and  $E = 48$  GPa, respectively. And the other two alloys with the cluster formulas of  $[(\text{Mo}_{0.5}\text{Sn}_{0.5}) - (\text{Ti}_{13}\text{Zr}_1)](\text{Nb}_{0.8}\text{Zr}_{0.2})$  (II-4) and  $[(\text{Mo}_{0.5}\text{Sn}_{0.5}) - (\text{Ti}_{13}\text{Zr}_1)](\text{Nb}_{0.5}\text{Zr}_{0.5})$  (II-5) are also predicted by setting the objective function of  $\text{Min } |E - E_s|$  with  $E_s = 55$  and  $E_s = 60$  GPa, respectively.

Figure 6a gives the XRD patterns of these five predicted alloys by the reverse design in ML, in which the II-2, II-4, and II-5 alloys possess a single BCC- $\beta$  structure. While for II-1 and II-3 alloys, several weak diffraction peaks of the  $\alpha''$  martensite appear on the  $\beta$  matrix. Although the  $\text{Mo}_{\text{eq}}$  values of these alloys surpass the critical lower limit ( $(\text{Mo}_{\text{eq}})_C = 11.8$  wt.%) of  $\beta$  stabilization, they are still in the vicinity of this limit, at which the  $\beta$  stability is susceptible to external conditions. Figure 7 gives the tensile engineering stress-strain curves of these alloys, from which the



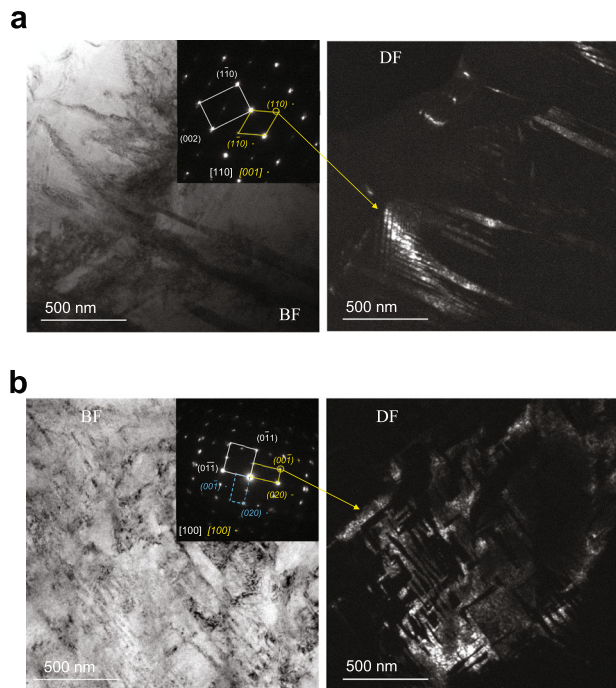
**Fig. 6 XRD patterns of the predicted alloys by ML in the reverse design. a Before tension. b After tension.**



$E$  values were measured, being  $E = 49 \pm 1$ ,  $46 \pm 1$ ,  $47 \pm 2$ ,  $58 \pm 2$ , and  $56 \pm 2$  GPa, respectively. Both the experimental and predicted  $E$  values of these designed alloys are all listed in Table 1. It could be found that the predicted  $E$  is consistent well with the experimental  $E$  for each alloy, which indicates that the cluster-formula-embedded ML can predict the alloy property (forward design) or alloy composition (reverse design) precisely. In addition, the  $Mo_{eq}$  values of these alloys, as well as alloy compositions (both at. % and wt. %) are also given in Table 1.

We also did the XRD analysis on the alloy samples after tension to check the variation of the crystalline structure induced by the applied load, which is shown in Fig. 6b. It is found that the diffraction peaks of the stress-induced  $\alpha''$  martensite appear obviously in the  $\beta$  matrix of each alloy, which is also validated by the double-yielding phenomenon in the engineering stress-strain curves shown in Fig. 7. It is ascribed to the stress-induced  $\alpha''$  martensite transformation when the  $\beta$  structural stability is relatively lower, which is consistent with the existing results<sup>20</sup>. The  $\beta$ -Ti alloys with such a transformation generally possess a relatively lower Young's modulus. Further detailed characterizations were carried on by TEM.

For the  $[(\text{Mo}_{0.3}\text{Sn}_{0.7}) - (\text{Ti}_{13.5}\text{Zr}_{0.5})](\text{Nb}_{1.5}\text{Ta}_{0.5})$  (II-2) and  $[(\text{Sn}_{0.5}\text{Zr}_{0.5}) - \text{Ti}_{14}](\text{Nb}_{2.5}\text{Ta}_{0.5})$  (II-3) alloys with the lowest Young's modulus ( $E = 46\text{--}47$  GPa), the SAED pattern shows that the II-2 alloy exhibits a single  $\beta$ -Ti structure without any other diffraction spots of either  $\alpha''$  or  $\omega$  phase, as presented in Fig. 8a. In fact, the Young's moduli of  $\alpha''$  and  $\omega$  phases are higher than that of  $\beta$ -Ti, especially the  $\omega$  phase<sup>20</sup>. While the diffraction spots of the  $\alpha''$  martensite appear in II-3 alloy, as found in Fig. 9a, in which both the bright-field and dark-field images demonstrate that a small amount of  $\alpha''$  martensitic plates are distributed in the  $\beta$ -Ti matrix. It is consistent with the XRD result, which might be resulted from the relatively lower  $M_{eq}$  value ( $M_{eq} = 11.95$  wt. %) in II-3 than that ( $M_{eq} = 13.20$  wt. %) in II-2. Nevertheless, there does not exist the  $\omega$  phase in II-3 alloy yet, in which the appearance of  $\omega$  could enhance the Young's modulus of  $\beta$ -Ti alloys drastically. After tension, the stress-induced  $\alpha''$  martensitic plates (including coarse and fine plates) appear in the  $\beta$  matrix in the II-2 alloy, as observed in the dark-field images (Fig. 8b). By comparison, a much larger amount of  $\alpha''$  martensitic plates appear in the tensioned II-3 alloy.



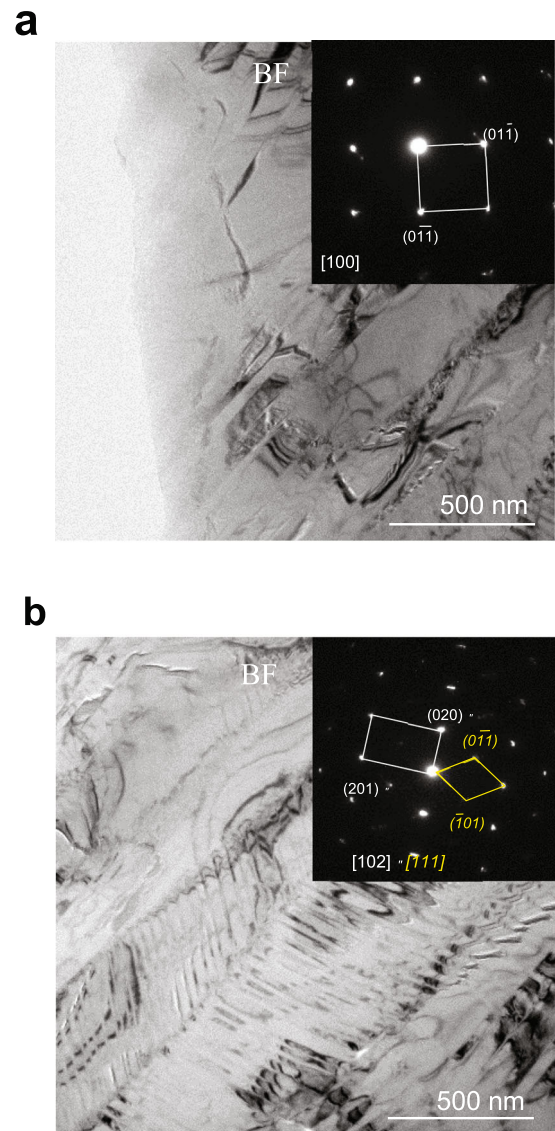
**Fig. 9** TEM morphological images and the corresponding SAED patterns of the  $[(\text{Sn}_{0.5}\text{Zr}_{0.5}) - \text{Ti}_{14}](\text{Nb}_{2.5}\text{Ta}_{0.5})$  (II-3) alloy at different states. **a** TEM results of the as-cast sample exhibiting a small amount of  $\alpha''$  martensitic plates in the  $\beta$ -Ti matrix. **b** TEM results after tension showing a much larger amount of  $\alpha''$  martensitic plates induced by applied stress in dark-field image.

as seen in Fig. 9b, which is closely related to the relatively lower  $\beta$  stability in the latter. Actually, it could be also confirmed by the fact that the double-yielding platform of II-3 is more obvious than that of II-2, indicating that a much higher amount of  $\alpha''$  martensite transformed from the  $\beta$  matrix during the tension process.

For the designed  $[(\text{Mo}_{0.5}\text{Sn}_{0.5}) - (\text{Ti}_{13}\text{Zr}_1)](\text{Nb}_{0.5}\text{Zr}_{0.5})$  (II-5) alloy with a given Young's modulus of  $E_5 = 60$  GPa, the TEM characterization (Fig. 10a) indicates that this alloys exhibit a single BCC  $\beta$ -Ti structure owing to the relatively higher  $Mo_{eq}$  value ( $Mo_{eq} = 13.53$  wt. %). After tension, there still exists a small amount of the stress-induced  $\alpha''$  martensite (Fig. 10b), similar to those in II-2 and II-3 alloys. All of these results certify again that the BCC- $\beta$  phase with lower stability would like to be accompanied by the protogenous  $\alpha''$  or stress-induced  $\alpha''$  martensite, which could render  $\beta$ -Ti alloys with lower Young's moduli.

## DISCUSSION

In the forward design of Loop I in ML, the relationship between the alloy composition ( $c_i$ ) and alloy property (Young's modulus,  $E$ ) was established by using the XGBoost model, in which the characteristic  $Mo_{eq}$  parameter for the representation the  $\beta$  structural stability of alloys is implemented to further restrict alloy compositions. Based on it, two alloys of I-1 and I-2 in Ti-Mo-Nb-Sn-Ta system were designed, but the experimental  $E$  values of these two alloys are somewhat different from the predicted ones, as seen in Table 1. This trend might be caused by the fact that alloy samples in this quinary system are not included in the ML database due to no existing experimental results, which would give rise to an uncertainty, and then reduce the prediction accuracy. More importantly, in the reverse design of Loop I, a given specific  $E$  value could predict many compositions because it is a multivariable function for a single target. For instance, if we set



**Fig. 10** TEM morphological images and the corresponding SAED patterns of the  $[(\text{Mo}_{0.5}\text{Sn}_{0.5}) - (\text{Ti}_{13}\text{Zr}_1)](\text{Nb}_{0.5}\text{Zr}_{0.5})$  (II-5) alloy at different states. **a** TEM results of the as-cast sample exhibiting a single  $\beta$ -Ti structure. **b** TEM results after tension showing a small amount of the stress-induced  $\alpha''$  martensite.

the specific Young's modulus as  $E_5 = 55$  GPa only with the  $Mo_{eq}$  constraint, we can obtain 85 compositions, in which the compositions vary with a step size of 1.0 at. %, as seen in Supplementary Table 2. It is difficult to validate all these predicted composition alloys by experiments, which could not manifest the superiority and efficiency of ML. In addition, there might exist some inevitable uncertainties that could not be reflected by the  $Mo_{eq}$  alone, which has been demonstrated by the present experiments for I-1 and I-2 alloys.

Then, we embedded the cluster-formula approach into the reverse design of ML (Loop II), which is a direct constraint to the composition. This approach considers the interactions among alloying elements with the base Ti, as well as among themselves. Our previous experimental results have demonstrated the validity of the cluster-formula approach, i.e., the  $\beta$ -Ti alloys with cluster formulas generally possess relatively lower Young's moduli, as evidenced by the quinary  $\beta$ -Ti  $[(\text{Mo}_{0.5}\text{Sn}_{0.5}) - (\text{Ti}_{13}\text{Zr}_1)](\text{Nb}_1)$  alloy



with  $E = 48$  GPa<sup>20</sup>. Thus, for a specific  $E$  value, there exist a limited number of alloy compositions or just only one composition in a given multicomponent system. For example, there have only two compositions (II-1, II-2) in the Ti–Mo–Nb–Zr–Sn–Ta hexa-component system and one composition (II-3) in the Ti–Nb–Zr–Sn–Ta quinary system to correspond to the  $Min E$  (Optimization problem given in Eq. (6)). More significantly, the experimental  $E$  values of these three composition alloys are in a great consistence with the predicted ones (Table 1). Furthermore, these two complex systems are not included in the ML database since no existing experimental results have been reported. Similar tendency also appears in II-4 and II-5 alloys for a specific  $E$  value. Therefore, by embedding the cluster formula and  $Mo_{eq}$  into the ML, we can optimize alloy compositions orientated by targets more effectively and precisely than the previous random research. In fact, when selecting alloy compositions in the forward design of ML, the compositions should satisfy the cluster-formula approach for a better prediction of  $E$ . It means that implementation of characteristic parameters that represent the features of a given system is crucial to establish the ML model, which is contributed to the prediction and design of the alloy composition and property with a much higher accuracy.

In our previous alloys, we applied the cluster-formula approach into other compositionally complex systems to explore the relationship between the alloy composition and properties<sup>37,64</sup>. For instance, it is found that the compositions of Ni-base single crystal superalloys with prominent creep resistance satisfy a uniform cluster formula of  $[\overline{Al} - \overline{Ni}_{12}](\overline{Al}_{1.5}\overline{Cr}_{1.5})$ , in which all the alloying elements are classified into three groups, Al series ( $\overline{Al}$ ), Cr series ( $\overline{Cr}$ ), and Ni series ( $\overline{Ni}$ )<sup>65</sup>. Since the creep rupture lifetime of these superalloys could be correlated to the composition in light of the cluster-formula approach, the cluster-formula-embedded ML model would predict the rupture lifetime of alloys and design new alloy compositions with high performance.

To summarize, a property-orientated alloy-design strategy combining the ML and characteristic parameters has been proposed to search for the low- $E$   $\beta$ -Ti alloys in the Ti–Mo–Nb–Zr–Sn–Ta system. The characteristic parameters are  $Mo_{eq}$  and cluster-composition formula, respectively. The XGBoost could establish the relationship between the alloy composition and Young's modulus  $E$  well in the forward design of ML with the guidance of  $Mo_{eq}$ . However, the experimental  $E$  value is indeed somewhat different from the predicted value of a given alloy when such alloy samples are not included into the ML database. On this basis, the cluster-formula approach was embedded in the reverse design to design new alloys with the desired  $E$ . Successfully, the experimental  $E$  values of predicted new alloys with the lowest  $E$  indeed reach the minimum ( $E = 46$ – $49$  GPa), which are  $[(Mo_{0.5}Sn_{0.5}) - (Ti_{1.3}Zr_{1.1})(Nb_{0.5}Ta_{0.5})]$ ,  $[(Mo_{0.3}Sn_{0.7}) - (Ti_{1.3.5}Zr_{0.5})(Nb_{1.5}Ta_{0.5})]$ , and  $[(Sn_{0.5}Zr_{0.5}) - Ti_{1.4}](Nb_{2.5}Ta_{0.5})$ , respectively, even though such kind of alloy samples are not included in the ML database. Therefore, for any specific  $E$ , this cluster-formula-embedded ML model could predict alloy composition precisely and efficiently in multicomponent systems. This design framework would be expected to predict other high-performance alloys, such as high-entropy alloys and Ni-base superalloys, in which the cluster-formula approach is of importance due to their chemical complexity.

## METHODS

### Modeling method

All the machine-learning models were created using the Scikit-learn (including RF and SVR) and XGBoost machine-learning libraries. The

reverse design was realized by the GA using geatpy library<sup>63</sup>, Scikit-learn, XGBoost, and geatpy are available under open-source licenses.

### Experimental procedures

The predicted alloys by ML were prepared by copper-mold suction cast processing, in which the experimental details are same as those in ref. <sup>19</sup>. Crystalline structures of alloys were first analyzed by the BRUKER D8 XRD with a Cu  $K\alpha$  radiation ( $\lambda = 0.15406$  nm). The microstructures were then verified with the JEM2100F FEG scanning transmission electron microscopy, in which the samples were prepared by twin-jet electro-polishing in a solution of 6%  $HClO_4 + 59\%$   $CH_3OH + 35\%$   $CH_3(CH_2)_3OH$  (volume fraction) at about 243 K. Tensile tests were executed on an UTM5504-G Material Test System (MTS) with a strain gage, where the tensile rate is set as 0.5 mm/min. The gauge size of rod tensile samples is  $3 \times 25$  mm (diameter  $\times$  length), and three samples for each composition alloy were measured to assure the reliabilities of tensile data.

### DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### CODE AVAILABILITY

The codes that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 9 January 2020; Accepted: 2 July 2020;

Published online: 21 July 2020

## REFERENCES

- Niinomi, M. Mechanical properties of biomedical titanium alloys. *Mater. Sci. Eng. A* **243**, 231–236 (1998).
- Long, M. & Rack, H. J. Titanium alloys in total joint replacement—a materials science perspective. *Biomaterials* **19**, 1621–1639 (1998).
- Hao, Y., Li, S., Sun, S., Zheng, C. & Yang, R. Elastic deformation behaviour of Ti-24Nb-4Zr-7.9Sn for biomedical applications. *Acta Biomater.* **3**, 277–286 (2007).
- Saito, T. et al. Multifunctional alloys obtained via a dislocation-free plastic deformation mechanism. *Science* **300**, 464–467 (2003).
- Ozaki, T., Matsumoto, H., Watanabe, S. & Hanada, S. Beta Ti alloys with low Young's modulus. *Mater. Trans.* **45**, 2776–2779 (2004).
- Wang, K. K., Gustavson, L. J. & Dumbleton, J. H. Microstructure and properties of a new beta titanium alloy, Ti-12Mo-6Zr-2Fe, developed for surgical implants. *Med. Appl. Titan. its Alloy*. **1272**, 76–87 (1996).
- Arciniegas, M., Manero, J. M., Peña, J., Gil, F. J. & Planell, J. A. Study of new multifunctional shape memory and low elastic modulus Ni-Free Ti alloys. *Mater. Trans. A* **39**, 742–751 (2008).
- Bania, P. J. Beta titanium alloys and their role in the titanium industry. *JOM-US* **46**, 16–19 (1994).
- Wang, Q., Dong, C. & Liaw, P. K. Structural stabilities of  $\beta$ -Ti alloys studied using a new Mo equivalent derived from  $[\beta/(\alpha + \beta)]$  phase-boundary slopes. *Mater. Trans. A* **46**, 3440–3447 (2015).
- Abdel-Hady, M., Hinoshita, K. & Morinaga, M. General approach to phase stability and elastic properties of  $\beta$ -type Ti-alloys using electronic parameters. *Scr. Mater.* **55**, 477–480 (2006).
- Kuroda, D., Niinomi, M., Morinaga, M., Kato, Y. & Yashiro, T. Design and mechanical properties of new  $\beta$  type titanium alloys for implant materials. *Mater. Sci. Eng. A* **243**, 244–249 (1998).
- Wang, Y., Dai, S., Chen, F., Yu, X. & Zhang, Y. Design, strength prediction of Ti35Nb5Sn7ZrMo alloys with low elastic moduli and experimental verification on their mechanical properties. *Rare Met.* **33**, 657–662 (2014).
- Bagariatskii, I. A., Nosova, G. I. & Tagunova, T. V. Factors in the formation of metastable phases in titanium-base alloys. *Sov. Phys. Dokl.* **3**, 1059–1064 (1958).
- Massalski, T. B. Comments concerning some features of phase diagrams and phase transformations. *Mater. Trans.* **51**, 583–596 (2010).
- Dong, C. et al. From clusters to phase diagrams: composition rules of quasi-crystals and bulk metallic glasses. *J. Phys. D.* **40**, 273–291 (2007).
- Wang, Z. et al. Composition design procedures of Ti-based bulk metallic glasses using the cluster-plus-glue-atom model. *Acta Mater.* **111**, 366–376 (2016).

17. Pang, C., Jiang, B., Shi, Y., Wang, Q. & Dong, C. Cluster-plus-glue-atom model and universal composition formulas [cluster](glue atom) $x$  for BCC solid solution alloys. *J. Alloy Compd.* **652**, 63–69 (2015).
18. Wang, Q., Ji, C., Wang, Y., Qiang, J. & Dong, C.  $\beta$ -Ti alloys with low Young's modulus interpreted by cluster-plus-glue-atom model. *Metall. Mater. Trans. A* **44**, 1872–1879 (2013).
19. Jiang, B. et al. Effects of Nb and Zr on structural stabilities of Ti-Mo-Sn-based alloys with low modulus. *Mater. Sci. Eng. A* **687**, 1–7 (2017).
20. Jiang, B. et al. Structural stability of the metastable  $\beta$ -[(Mo<sub>0.5</sub>Sn<sub>0.5</sub>)-(Ti<sub>13</sub>Zr<sub>1</sub>)]Nb<sub>1</sub> alloy with low Young's modulus at different states. *Metall. Mater. Trans. A* **48**, 3912–3919 (2017).
21. Xin, Y., Ma, T., Ju, X. & Qiu, J. In-situ observation of point defect and precipitate involvement of CLAM steel under electron irradiation. *J. Mater. Sci. Technol.* **29**, 467–470 (2013).
22. Zhou, X., Liu, C., Yu, L., Liu, Y. & Li, H. Phase transformation behavior and microstructural control of high-Cr martensitic/ferritic heat-resistant steels for power and nuclear plants: A review. *J. Mater. Sci. Technol.* **31**, 235–242 (2015).
23. Ramprasath, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A. & Kim, C. Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* **3**, 54 (2017).
24. Raccuglia, P. et al. Machine-learning-assisted materials discovery using failed experiments. *Nature* **533**, 73–76 (2016).
25. Deng, Z. et al. Machine learning aided study of sintered density in Cu-Al alloy. *Comput. Mater. Sci.* **155**, 48–54 (2018).
26. Xue, D. et al. An informatics approach to transformation temperatures of NiTi-based shape memory alloys. *Acta Mater.* **125**, 532–541 (2017).
27. Wang, C. et al. A property-oriented design strategy for high performance copper alloys via machine learning. *npj Comput. Mater.* **5**, 87 (2019).
28. Huang, W., Martin, P. & Zhuang, L. Machine-learning phase prediction of high-entropy alloys. *Acta Mater.* **169**, 225–236 (2019).
29. Wen, C. et al. Machine learning assisted design of high entropy alloys with desired property. *Acta Mater.* **170**, 109–117 (2019).
30. Kim, G. et al. First-principles and machine learning predictions of elasticity in severely lattice-distorted high-entropy alloys with experimental validation. *Acta Mater.* **181**, 124–138 (2019).
31. Shen, C. et al. Physical metallurgy-guided machine learning and artificial intelligent design of ultrahigh-strength stainless steel. *Acta Mater.* **179**, 201–214 (2019).
32. Chen, T. & Guestrin, C. Xgboost: A Scalable Tree Boosting System. In *Proc. 22nd ACM Sigkdd International Conference on Knowledge Discovery And Data Mining*, 785–794 (ACM, 2016).
33. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
34. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
35. Okazaki, Y. & Gotoh, E. Comparison of metal release from various metallic bio-materials in vitro. *Biomaterials* **26**, 11–21 (2005).
36. Takeuchi, A. & Inoue, A. Classification of bulk metallic glasses by atomic size difference, heat of mixing and period of constituent elements and its application to characterization of the main alloying element. *Mater. Trans.* **46**, 2817–2829 (2005).
37. Jiang, B., Wang, Q., Dong, C. & Liaw, P. K. Exploration of phase structure evolution induced by alloying elements in Ti alloys via a chemical-short-range-order cluster model. *Sci. Rep.* **9**, 3404 (2019).
38. You, L. & Song, X. A study of low Young's modulus Ti-Nb-Zr alloys using  $d$  electrons alloy theory. *Scr. Mater.* **67**, 57–60 (2012).
39. Matsumoto, H., Watanabe, S. & Hanada, S. Beta TiNbSn alloys with low Young's modulus and high strength. *Mater. Trans.* **46**, 1070–1078 (2005).
40. Zhang, D. C. et al. Effect of Sn addition on the microstructure and superelasticity in Ti-Nb-Mo-Sn Alloys. *J. Mech. Behav. Biomed.* **13**, 156–165 (2012).
41. Hao, Y. L., Li, S. J., Sun, S. Y. & Yang, R. Effect of Zr and Sn on Young's modulus and superelasticity of Ti-Nb-based alloys. *Mater. Sci. Eng. A* **441**, 112–118 (2006).
42. Meng, Q., Liu, Q., Guo, S., Zhu, Y. & Zhao, X. Effect of thermo-mechanical treatment on mechanical and elastic properties of Ti-36Nb-5Zr alloy. *Prog. Nat. Sci. Mater. Int.* **25**, 229–235 (2015).
43. Sakaguchi, N., Niinomi, M., Akahori, T., Saito, T. & Furuta, T. Effects of alloying elements on elastic modulus of Ti-Nb-Ta-Zr system alloy for biomedical applications. *Mater. Sci. Forum* **449–452**, 1269–1272 (2004).
44. Moraes, P. E. L., Contieri, R. J., Lopes, E. S. N., Robin, A. & Caram, R. Effects of Sn addition on the microstructure, mechanical properties and corrosion behavior of Ti-Nb-Sn alloys. *Mater. Charact.* **96**, 273–281 (2014).
45. Kent, D., Wang, G. & Dargusch, M. Effects of phase stability and processing on the mechanical properties of Ti-Nb based  $\beta$  Ti alloys. *J. Mech. Behav. Biomed.* **28**, 15–25 (2013).
46. Wei, Q. et al. Influence of oxygen content on microstructure and mechanical properties of Ti-Nb-Ta-Zr alloy. *Mater. Des.* **32**, 2934–2939 (2011).
47. Yu, Z., Zhou, L., Luo, L., Fan, M. & Fu, Y. Investigation on mechanical compatibility matching for biomedical titanium alloys. *Key Eng. Mater.* **288–289**, 595–598 (2005).
48. Yu, Z., Zhou, L., Fan, M. & Yuan, S. Investigation on near- $\beta$  titanium alloy Ti-5Zr-3Sn-5Mo-15Nb for surgical implant materials. *Mater. Sci. Forum* **475**, 2353–2358 (2005).
49. Kathy, W. The use of titanium for medical applications in the USA. *Mater. Sci. Eng. A* **213**, 134–137 (1996).
50. Raducanu, D. et al. Mechanical and corrosion resistance of a new nanostructured Ti-Zr-Ta-Nb alloy. *J. Mech. Behav. Biomed.* **4**, 1421–1430 (2011).
51. Laheurte, P. et al. Mechanical properties of low modulus  $\beta$  titanium alloys designed from the electronic approach. *J. Mech. Behav. Biomed.* **3**, 565–573 (2010).
52. Elias, L. M., Schneider, S. G., Schneider, S., Silva, H. M. & Malvisi, F. Microstructural and mechanical characterization of biomedical Ti-Nb-Zr-Ta alloys. *Mater. Sci. Eng. A* **432**, 108–112 (2006).
53. Tang, X., Ahmed, T. & Rack, H. J. Phase transformations in Ti-Nb-Ta and Ti-Nb-Ta-Zr alloys. *J. Mater. Sci.* **35**, 1805–1811 (2000).
54. Williams, J. C., Froes, F. H. & Yoltou, C. F. Some observations on the structure of Ti-11.5 Mo-6 Zr-4.5 Sn (Beta III) as affected by processing history. *Metall. Trans. A* **11**, 356–358 (1980).
55. Hsu, H., Wu, S., Hsu, S., Kao, W. & Ho, W. Structure and mechanical properties of as-cast Ti-5Nb-based alloy with Mo addition. *Mater. Sci. Eng. A* **579**, 86–91 (2013).
56. Correa, D., Kuroda, P. & Grandini, C. Structure, microstructure, and selected mechanical properties of Ti-Zr-Mo alloys for biomedical applications. *Adv. Mater. Res.* **922**, 75–80 (2014).
57. Bertrand, E. et al. Synthesis and characterisation of a new superelastic Ti-25Ta-25Nb biomedical alloy. *J. Mech. Behav. Biomed.* **3**, 559–564 (2010).
58. Gordin, D. M. et al. Synthesis, structure and electrochemical behavior of a beta Ti-12Mo-5Ta alloy as new biomaterial. *Mater. Lett.* **59**, 2936–2941 (2005).
59. Hsu, H., Wu, S., Hsu, S., Syu, J. & Ho, W. The structure and mechanical properties of as-cast Ti-25Nb-xSn alloys for biomedical applications. *Mater. Sci. Eng. A* **568**, 1–7 (2013).
60. Xu, W. et al. A new ultrahigh-strength stainless steel strengthened by various coexisting nanoprecipitates. *Acta Mater.* **58**, 4067–4075 (2010).
61. Xu, W., Castillo, P. E. J. R. & Zwaag, S. V. D. A combined optimization of alloy composition and aging temperature in designing new UHS precipitation hardenable stainless steels. *Comp. Mater. Sci.* **45**, 467–473 (2009).
62. Lu, Q., Xu, W. & van der Zwaag, S. A strain-based computational design of creep-resistant steels. *Acta Mater.* **64**, 133–143 (2014).
63. Jazbin, et al. geatpy: the genetic and evolutionary algorithm toolbox with high performance in python. <http://www.geatpy.com/> (2020).
64. Ma, Y. et al. Controlled formation of coherent cuboidal nanoprecipitates in body-centered cubic high-entropy alloys based on Al<sub>2</sub>(Ni,Co,Fe,Cr)<sub>14</sub> compositions. *Acta Mater.* **147**, 213–225 (2018).
65. Zhang, Y. et al. High-Temperature Structural Stabilities of Ni-based single-crystal superalloys Ni-Co-Cr-Mo-W-Al-Ti-Ta with varying Co contents. *Acta Metall. Sin.* **31**, 127–133 (2018).

## ACKNOWLEDGEMENTS

It was supported by the National Natural Science Foundation of China [No. 91860108 and U1867201], the National Key Research and Development Plan (2017YFB0702401), National Science Foundation of Liaoning Province of China (Grant No. 2019-KF-05-01), and the Fundamental Research Funds for the Central Universities (DUT19LAB01).

## AUTHOR CONTRIBUTIONS

F.Y. did all of the experimental work. Z.L. carried out the machine learning. Q.W. and C.D. supervised the project and conceived the idea. B.B.J. collected the data. B.J.Y., P.C.Z., W.X., and P.K.L. analyzed the results. All authors discussed the results and contributed to writing of the paper.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41524-020-00372-w>.

**Correspondence** and requests for materials should be addressed to Z.L. or Q.W.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020