materials

MDPI

Article

Composition Design of a Novel High-Temperature Titanium Alloy Based on Data Augmentation Machine Learning

Xinpeng Fu [1,2], Boya Li [3,*], Binguo Fu [1,2,*], Tianshun Dong [1,2] and Jingkun Li [1,2]

[1] State Key Laboratory of High Performance Roll Materials and Composite Forming, School of Materials Science and Engineering, Hebei University of Technology, Tianjin 300401, China; fuxp_1995@163.com (X.F.); dongtianshun111@163.com (T.D.); tjjk_zy@126.com (J.L.)
[2] Hebei Key Laboratory of New Functional Materials, School of Materials Science and Engineering, Hebei University of Technology, Tianjin 300401, China
[3] School of Materials and Chemical Technology, Institute of Science Tokyo, Tokyo 152-8552, Japan
* Correspondence: li.b.ah@m.titech.ac.jp (B.L.); fubinguohao@163.com (B.F.)

Abstract

The application fields of high-temperature titanium alloys are mainly concentrated in the aerospace, defense and military industries, such as the high-temperature parts of rocket and aircraft engines, missile cases, tail rudders, etc., which can greatly reduce the weight of aircraft while resisting high temperatures. However, traditional high-temperature titanium alloys containing multiple types of elements (more than six) have a complex impact on the solidification, deformation, and phase transformation processes of the alloys, which greatly increases the difficulty of casting and deformation manufacturing of aerospace and military components. Therefore, developing low-component high-temperature titanium alloys suitable for hot processing and forming is urgent. This study used data augmentation (Gaussian noise) to expedite the development of a novel quinary high-temperature titanium alloy. Utilizing data augmentation, the generalization abilities of four machine learning models (XGBoost, RF, AdaBoost, Lasso) were effectively improved, with the XGBoost model demonstrating superior prediction accuracy (with an R^2 value of 0.94, an RMSE of 53.31, and an MAE of 42.93 in the test set). Based on this model, a novel Ti-7.2Al-1.8Mo-2.0Nb-0.4Si (wt.%) alloy was designed and experimentally validated. The UTS of the alloy at 600 °C was 629 MPa, closely aligning with the value (649 MPa) predicted by the model, with an error of 3.2%. Compared to as-cast Ti1100 and Ti6242S alloy (both containing six elements), the novel quinary alloy has considerable high-temperature (600 °C) mechanical properties and fewer components. The microstructure analysis revealed that the designed alloy was an α+β type alloy, featuring a typical Widmanstätten structure. The fracture form of the alloy was a mixture of brittle and ductile fracture at both room and high temperatures.

Keywords: high-temperature titanium alloy; machine learning; data augmentation; microstructure; mechanical property

check for updates

Academic Editor: Murali Mohan Cheepu

Received: 8 May 2025
Revised: 26 June 2025
Accepted: 27 June 2025
Published: 30 June 2025

Citation: Fu, X.; Li, B.; Fu, B.; Dong, T.; Li, J. Composition Design of a Novel High-Temperature Titanium Alloy Based on Data Augmentation Machine Learning. *Materials* **2025**, *18*, 3099. https://doi.org/10.3390/ma18133099

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Titanium alloys are widely used in fields such as aerospace, medicine, defense and military industry due to their low density, high specific strength, and excellent corrosion resistance. The rapid development of the aerospace industry has spotlighted the requirement for advanced high-temperature titanium alloys [1–3]. Currently, the more typical high-temperature titanium alloys at 600 °C are IMI834 and Ti1100, both of which belong to

the Ti-Al-Sn-Zr-Mo-Si alloy system [4–6]. With the development of aerospace technology, titanium alloys need to withstand higher temperatures (more than 600 °C) on the basis of structural weight reduction. At the same time, titanium alloy components are increasingly showing trends of large size, thin-walled surfaces, variable thickness, and overall structural development, which poses new challenges for the processing and manufacturing of titanium alloys. However, there is a lack of systematic research on the composition design of high-temperature titanium alloys that combine heat resistance and formability both domestically and internationally. The development of traditional high-temperature titanium alloys is achieved by the addition of rare earth elements (Y, Er, Nd) or β-stabilizing elements (Nb, Ta, W) on the basis of the above-mentioned system, and performance is improved through alloying [7–10]. However, the addition of too many components has a complex impact on the solidification, deformation, and phase transformation processes of the alloys, which greatly increases the difficulty of material casting and deformation manufacturing. This approach also faces diminishing returns in performance improvement, increasing material costs, and recycling complexities [11,12]. Against this background, this study aimed to break the traditional system of high-temperature titanium alloys (Ti-Al-Sn-Zr-Mo-Si) and develop a novel high-temperature titanium alloy that contained only five elements. By optimizing the types and contents of alloying elements, the costs and dependence on scarce resources can be reduced without affecting performance.

The selection of elements underwent rigorous analysis and screening. The Al element is the most commonly used element in high-temperature titanium alloys. The addition of the Al element can enhance the thermal stability of titanium alloys and reduce the alloy density. Moreover, due to its strong strengthening effect and low cost, the Al element is usually used for manufacturing low-cost titanium alloys [13,14]. The Mo element can effectively enhance the strength of the alloy at room and high temperatures by strengthening the β-phase [15,16]. The Nb element is a relatively weak β-stabilizing element, which has an electronegativity similar to that of Ti and exhibits a high solid solubility in the α-phase. As such, it can effectively exert solid solution strengthening effects on both the α-phase and the β-phase simultaneously [17–19]. The Si element is also a commonly added element in high-temperature titanium alloys, which contributes to solid solution strengthening and forms silicides that pin dislocations and boundaries, further elevating alloy strength [20–22]. Meanwhile, the formation of Ti_5Si_3 silicide with a $D8_8$ hexagonal structure (lattice parameters a = 0.7444 nm and c = 0.5143 nm) features a high melting point (2130 °C) and low density (4.32 g/cm^3). Its thermal expansion coefficient is similar to that of titanium, thereby improving the creep resistance of the alloy [23,24]. Therefore, Al, Mo, Nb and Si were selected as the constituent elements of the alloy in this study.

The empirical trial-and-error method has the disadvantages of high cost and long material design cycle. For instance, exploring just 10 content levels across the four elements Al, Mo, Nb, and Si would generate 10,000 potential combinations, which severely restricts the development of novel alloys. With the rapid development of artificial intelligence (AI) technology and the accumulation of experimental data, the machine learning (ML) method has shown great potential in material research and development [25–27]. Liu et al. established four ML models (Adaptive Boosting, Light Gradient Boosting Machine, Voting, and Stacking) for the design of metastable β-titanium alloys, among which the Light Gradient Boosting Machine model had the optimal prediction accuracy. Based on this model, a metastable β-titanium alloy (Ti-5.5Cr-5Al-4Mo-3Nb-2Zr) was designed, and the experimental values of ultimate tensile strength (UTS), yield strength (YS), and elongation (EL) were in good agreement with the values predicted by the model [28]. Wu et al. used an Artificial Neural Network (ANN) model to design a Ti-12Nb-12Zr-12Sn alloy. This alloy met the requirements of biocompatibility, low modulus, and low cost, and was expected to

be used in orthopedics and prosthetic implantation [29]. Oh et al. constructed a dataset of Mn content, solution treatment temperature, UTS, and EL to train an ANN model. Based on this model, a low-cost TRIP titanium alloy, Ti-4Al-2Fe-1.4Mn, was developed. After solution treatment at 883 °C, this alloy had an ultra-high specific strength (289 MPa·cm^3/g) and excellent elongation (34%) [30]. However, the potential of ML in high-temperature titanium alloy design remains untapped.

The purpose of this study is to find a design method that combines low-component and high-heat-resistant temperature titanium alloys. This study established a dataset of the alloy composition, test temperature, and UTS of high-temperature titanium alloys, and data augmentation was carried out on the original dataset. The predictive performance of four ML models was systematically evaluated, and an optimal ML model was used to develop a novel high-temperature titanium alloy. Beyond high predictive accuracy, the model's efficacy was validated through alloy preparation and performance testing, providing a new idea for the development of novel high-temperature titanium alloys. The novel titanium alloy successfully reduces the number of alloying elements while ensuring high performance. And this is of great significance for promoting the application of high-temperature titanium alloy castings in the aerospace industry.

2. Material Design and Experiment

2.1. Alloy Design Framework

Figure 1 schematically illustrates the design process of a novel high-temperature titanium alloy through ML methods. As shown in Figure 1, the first step is data collection and augmentation: through a literature review, an original dataset encompassing alloy compositions, testing temperatures, and UTS was constructed, and an augmented dataset was obtained through data augmentation. The second step is ML model selection: the generalization ability and prediction accuracy of four ML models were compared to determine the optimal ML model. The third step is alloy design: the physical metallurgy model was combined with the genetic algorithm for composition design. The fourth step is experimental verification: the properties and microstructure of the alloy were analyzed to verify the effectiveness of the design scheme in this study.

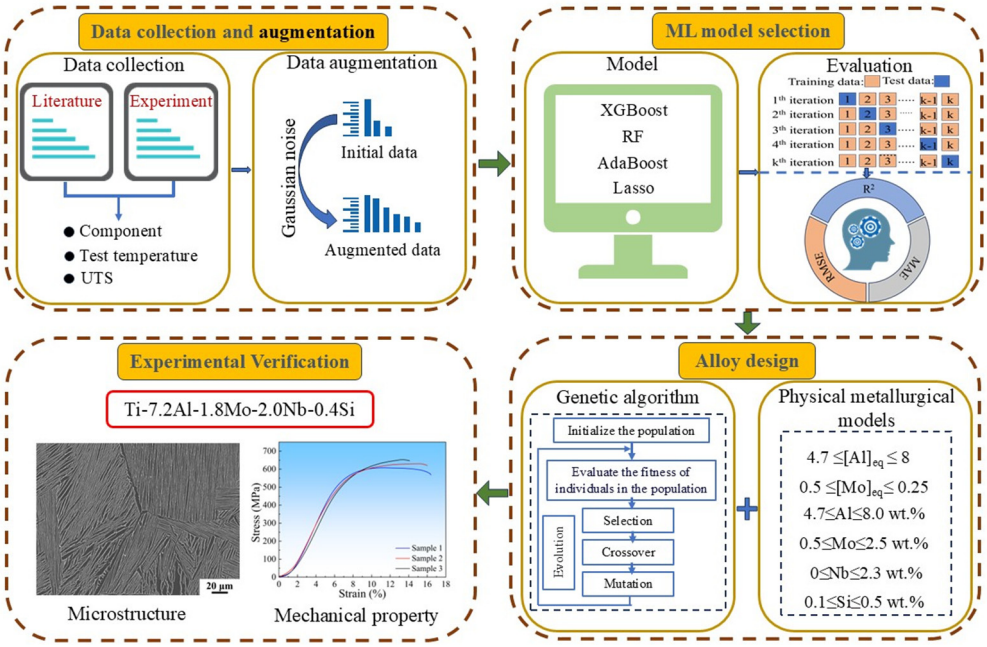

Figure 1. ML design scheme for novel high-temperature titanium alloy.

2.2. Dataset Establishment

In this study, a dataset containing 114 samples was established through consulting the literature, as shown in Table 1. The dataset consists of three parts, namely, the alloy composition and testing temperature as input features, and the UTS as the output feature. The alloy composition includes ten elements, Al, Sn, Zr, Mo, Si, Nb, Ta, W, Y, and V. The testing temperature ranges from 20 °C to 750 °C, and the UTS ranges from 381 MPa to 1328 MPa. In order to ensure the scientific rigor and applicability of the data, the following data screening criteria were set: (1) The alloy system was limited to near-α type or α+β type titanium alloys with high-temperature application potential. (2) All alloys involved in the data were prepared by casting forming. (3) The collected values of the UTS were all the properties of the alloys in the as-cast condition, excluding the interference of subsequent thermomechanical treatments (plastic deformation, heat treatment).

Table 1. Dataset information.

Variable		Unit	Data Range	
Input	Composition	wt.%	Al	3.2~10.0
			Sn	0~6.2
			Zr	0~11.0
			Mo	0~4.05
			Si	0~1.5
			Nb	0~6.5
			Ta	0~4.0
			W	0~3.6
			Y	0~0.62
			V	0~4.5
	Testing temperature	°C	T	20~750
Output	Ultimate tensile strength	MPa	UTS	381~1328

2.3. Data Augmentation

To further enhance the data quality, this study attempted to achieve data augmentation by adding noise points to the original data [31]. This approach is particularly relevant to material science because it simulates the uncontrollable errors in practical alloy design. For example, during the smelting of the alloy, the loss of raw materials can lead to a deviation between the actual composition and the nominal composition. In the application of material design, Ye et al. carried out data augmentation through the Gaussian noise method, achieving accurate prediction of the phases and hardness of high-entropy alloys, which demonstrated its effectiveness [32]. The enhancement of data with Gaussian noise can be achieved through Equation (1) [33]:

$$x_{aug}^{i} = x + \sigma \cdot \varepsilon^{(i)}, i = 1, 2, \ldots\ldots, n \quad (1)$$

where x_{aug}^{i} is the ith augmented data point, x is the original data, σ is the standard deviation of the Gaussian noise, $\varepsilon^{(i)}$~N (0, 1) indicates that the randomly sampled variable follows a normal distribution, and n is the number of augmented data points generated for each original data point. In this study, σ and n were set to 0.01 and 1, respectively.

2.4. Model Establishment and Evaluation

According to the "No Free Lunch Theorem" proposed by Wolpert, there is no single algorithm that can consistently outperform all other algorithms across all possible problems or datasets [34]. Therefore, in this study, four different ML models, namely, eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Adaptive Boosting (AdaBoost), and

Least absolute shrinkage and selection operator (Lasso), were established. Firstly, the impact of data augmentation on the performance of the ML model was evaluated through 10-fold cross-validation. Figure 2 is a schematic diagram of 10-fold cross-validation. Its principle is as follows: the dataset is divided into 10 non-overlapping subsets. In each iteration, 9 of these subsets are selected as the training set to train the model, and the remaining 1 subset is used as the test set to evaluate the performance of the model. This process is repeated 10 times to ensure that each subset is used as the test set. Finally, the average value of the coefficient of determination (R^2) of the 10 validations is calculated to obtain the 10-fold cross-validation score. The closer the score is to 1, the better the generalization ability of the model.

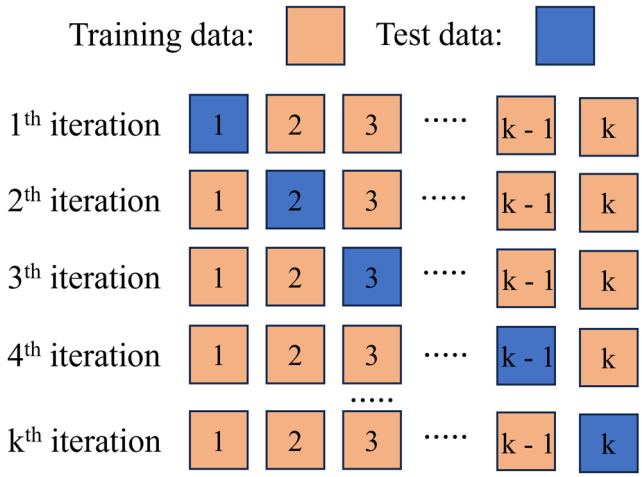

Figure 2. Schematic diagram of 10-fold cross-validation.

Next, the four ML models were trained and tested to determine the optimal model. The ratio of the test set to the training set was 1:4. The prediction accuracy of the four ML models was evaluated using the root mean squared error (RMSE), the mean absolute error (MAE), and the coefficient of determination (R^2). The equation is as follows:

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(y_i - f(x_i))| \quad (3)$$

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

where n is the number of samples in the dataset, $\overline{y}$ is the average value of all experimental values in the dataset, and y_i and $f(x_i)$ are the experimental value of each sample in the dataset and the corresponding predicted value of the ML model, respectively.

2.5. Alloy Design

The content ranges of four elements, Al, Mo, Nb and Si, were determined by combining physical metallurgy models such as Aluminum equivalent ($[Al]_{eq}$) and Molybdenum equivalent ($[Mo]_{eq}$) with domain knowledge. By imposing compositional constraints, the design efficiency of the ML model can be improved, and the reliability of the design results can be ensured [29,35]. The genetic algorithm was employed for composition screening, and the optimal alloy composition along with its corresponding predicted value of the UTS was obtained. The parameter settings of the genetic algorithm were as follows: the

population size was set to 50, the number of generations was set to 1000, the crossover probability was set to 0.8, and the mutation probability was set to 0.2.

2.6. Experiment

The alloy was prepared by using a vacuum suspension melting furnace equipped with a water-cooled copper crucible, and the ultimate vacuum degree of the melting furnace was 10^{-2} Pa. Pure metals of Ti (99.90 wt.%), Al (99.99 wt.%), Mo (99.95 wt.%), Nb (99.95 wt.%), and Si (99.95 wt.%) were used as raw materials. To ensure the uniformity of the composition, the alloy ingot was melted three times. Alloy phases were examined by an X-ray diffractometer (XRD) with Cu Kα X-radiation (Rigaku Smartlab 9 kW, Akishima, Japan). The XRD sample was scanned over a range of 20°~80° with a scanning speed of 10°/min. The microstructure and tensile fracture surface of the alloy were observed using a scanning electron microscope (SEM, Hitachi SU3800, Hitachinaka, Japan). The microstructure sample was chemically etched using Kroll reagent ($HF:HNO_3:H_2O$ = 1:2:7, vol.%). The room-temperature tensile property of as-cast alloy was tested using an Instron-5982 (Boston, MA, USA) universal electronic testing machine with a tensile rate of 1 mm/min. The tensile property of as-cast alloy at 600 °C was tested using a CMT-5205 (Ningbo, China) electronic universal testing machine with a tensile rate of 0.6 mm/min and a holding time of 10 min according to the GB/T 228.2-2015 standard [36]. According to the ASTM: E8/E8M-16 standard [37], a 15 mm gauge length extensometer was used to measure the strain to exclude the influence of the elastic strain of the grips. The yield strengths were calculated using the 0.2% strain offset method, as described in [38]. To ensure the reliability and repeatability of the data, three specimens were tested under each condition. The dimensions of the tensile specimens are shown in Figure 3.

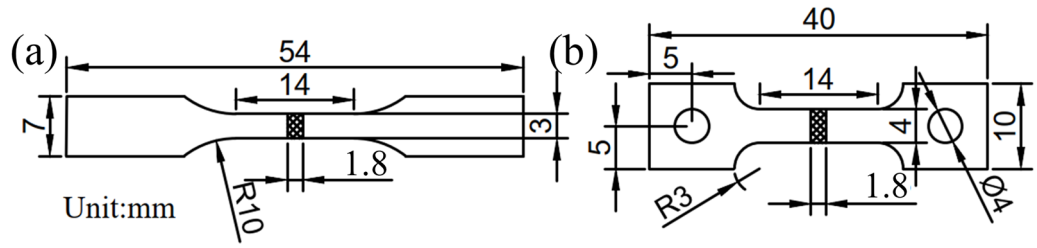

Figure 3. Dimensions of the alloy tensile specimens: (**a**) room temperature; (**b**) 600 °C.

3. Results and Discussion

3.1. Gaussian Enhancement

Figure 4 schematically illustrates the process of introducing Gaussian noise into the original dataset. The data in three different colors represent the alloy composition, testing temperature, and UTS, respectively. The data in the first five columns show that by introducing Gaussian noise into the original data values, the so-called "actual values" or noisy samples could be generated (the complete original data and augmented data can be found in Supplementary Tables S1 and S2). This method effectively doubles the size of the dataset. In order to evaluate the impact of data augmentation on the performance of the ML models, the performance of the four ML models was evaluated using 10-fold cross-validation. As shown in Figure 5, after data augmentation, the 10-fold cross-validation scores of the XGBoost, RF, AdaBoost, and Lasso models increased from 0.87, 0.84, 0.82, and 0.83 to 0.92, 0.88, 0.88, and 0.86, respectively. This enhancement can be attributed to the increased data diversity by data augmentation. The expanded dataset enabled the models to learn more features during the training process, thereby enhancing the generalization ability of the model.

Al/wt.%	Sn/wt.%	Zr/wt.%	Mo/wt.%	Si/wt.%	Nb/wt.%	Ta/wt.%	W/wt.%	Y/wt.%	V/wt.%	T/°C	UTS/MPa
5.61	3.69	3.45	0.99	0.38	0.40	0.20	0	0	0	20	940
5.91	4.07	3.92	0.51	0.27	0.82	0	0	0	0	550	645
5.92	2.12	3.88	2.03	0.10	0	0	0	0	0	600	571
5.41	2.52	5.33	0.68	0.19	1.35	1.33	1.35	0.62	0	650	603
6.52	4.25	5.50	2.73	0.40	1.77	0	3.58	0.10	0	700	661

Original data

Set Gaussian noise:n=1,σ=0.01

Augmented data

Al/wt.%	Sn/wt.%	Zr/wt.%	Mo/wt.%	Si/wt.%	Nb/wt.%	Ta/wt.%	W/wt.%	Y/wt.%	V/wt.%	T/°C	UTS/MPa
5.61	3.67	3.44	0.97	0.37	0.40	0.19	0	0	0	19.9	940
5.92	4.06	3.90	0.50	0.26	0.79	0	0	0	0	549.9	645
5.93	2.13	3.89	2.02	0.09	0	0	0	0	0	599.9	571
5.41	2.51	5.33	0.67	0.20	1.35	1.33	1.34	0.61	0	649.9	603
6.53	4.25	5.50	2.72	0.38	1.76	0	3.58	0.12	0	699.9	661

Figure 4. Schematic diagram of adding Gaussian noise to original data.

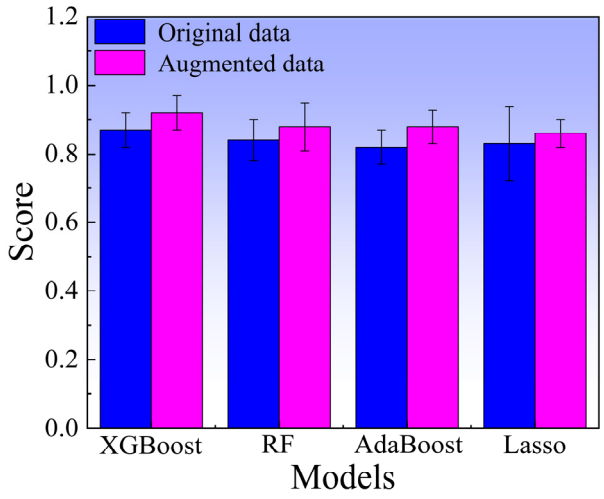

Figure 5. Comparison between the original dataset and the augmented dataset.

3.2. Selection of Optimal ML Model

The optimal model was determined by comparing the prediction performances of the four ML models on the augmented dataset. Figure 6 shows the prediction results of the UTS by four ML models. The XGBoost model exhibits the most concentrated data distributions, achieving R^2 values of 0.99 and 0.94 for the training and test sets, respectively, which are substantially higher than those of the other models. Figure 7 shows the RMSE and MAE of four ML models in the task of predicting the UTS. The results indicate that the XGBoost model has the smallest prediction error. The RMSE and MAE values of the training set are 21.32 and 15.96, respectively, while those of the test set are 53.31 and 42.93, respectively. In conclusion, the XGBoost model demonstrates the best prediction performance in the task of predicting the UTS. It is capable of accurately establishing the mapping relationship between the input variables (alloy composition, test temperature) and the output variable (UTS). XGBoost's exceptional performance stems from integrating a regularization term into its objective function, which controls model complexity and prevents overfitting. Additionally, XGBoost gradually optimizes the objective function through gradient boosting and combines second-order Taylor expansion to accurately

fit non-linear relationships, enabling it to perform outstandingly in complex non-linear problems [39–41]. Compared with RF, AdaBoost, and Lasso, the XGBoost model is more suitable for predicting the non-linear functional relationship between the UTS of alloys and multi-dimensional independent variables (alloy composition and test temperature). Therefore, the XGBoost model was selected for the design of high-temperature titanium alloy compositions in this study.

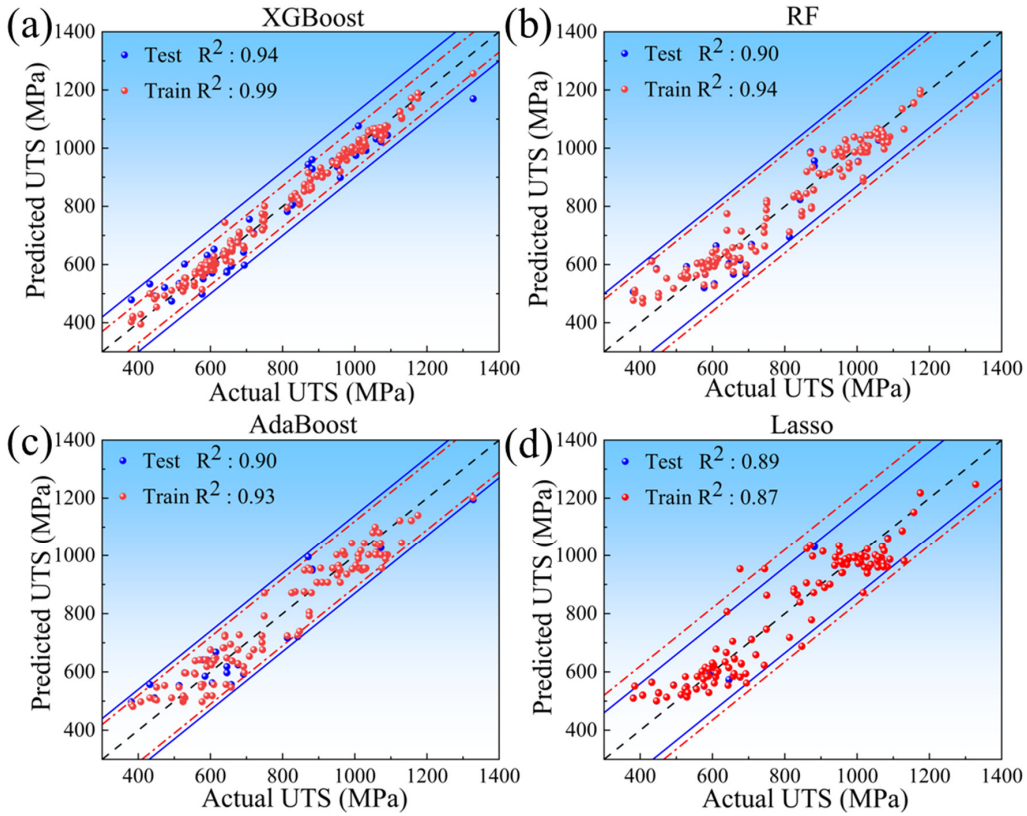

Figure 6. The prediction results of the UTS by four ML models: (**a**) XGBoost; (**b**) RF; (**c**) Adaboost; (**d**) Lasso. (the black dashed line represents the perfect prediction line, the red dot-dashed line represents the distribution range of the training set, and the blue solid line represents the distribution range of the test set).

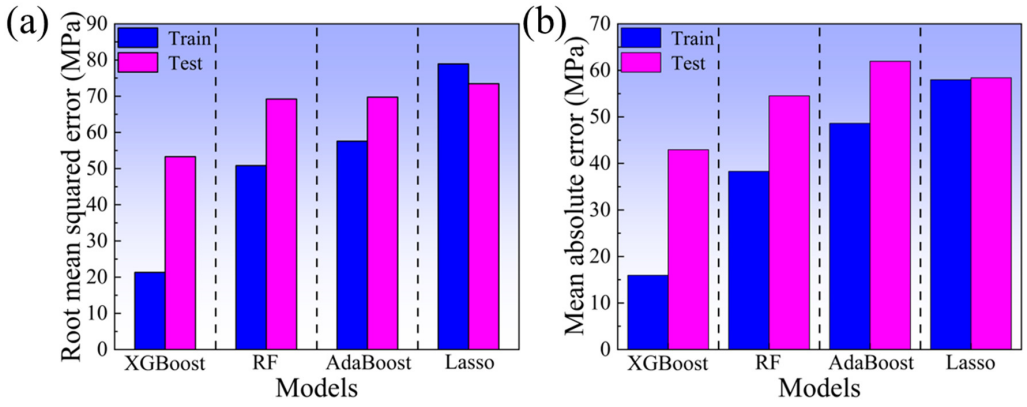

Figure 7. Comparison chart of RMSE and MAE of four ML models: (**a**) RMSE; (**b**) MAE.

3.3. Optimal Alloy Composition Design

When using ML for alloy composition design, if there are no constraints on the alloy composition, it is possible to generate a large number of solutions that are "high

performance based on AI hallucinations" but not feasible in practice. When the content of the Al element in the alloy is relatively high, a large amount of the brittle phase (Ti_3Al) will precipitate from the α phase under high-temperature service conditions, thus reducing the plasticity of the alloy [42,43]. Meanwhile, elements such as Mo, Nb, and Si are β-stabilizing elements. Excessive addition of these elements will transform the alloy into a β-type titanium alloy, failing to meet the requirements for high-temperature titanium alloys. To address these challenges, this study utilized the physical metallurgical model $[Al]_{eq}$ to constrain the Al content, and the physical metallurgical model $[Mo]_{eq}$ to constrain the Mo and Nb contents, with their expressions shown in Equations (5) and (6) [44,45], respectively. Figure 8 shows the $[Al]_{eq}$ and $[Mo]_{eq}$ of the typical high-temperature titanium alloys, and the results show that the distribution interval of $[Al]_{eq}$ is 4.7~8.0, and the distribution interval of $[Mo]_{eq}$ is 0.5~3.2. Accordingly, the content of Al was limited to 4.7~8.0 wt.%, Mo was 0.5~2.5 wt.%, and Nb was 0~2.3 wt.%. For Si, the content in commonly used high-temperature titanium alloys is 0.1~0.5 wt.% [46]. Next, the genetic algorithm was used to search within the defined compositional range, targeting the UTS at 600 °C. The alloy composition corresponding to the highest UTS was calculated. Table 2 shows the alloy composition and its predicted UTS value at 600 °C.

$$[Al]_{eq} = Al\% + (Sn/3)\% + (Zr/6)\% + 10(O)\% \quad (5)$$

$$[Mo]_{eq} = Mo\% + (Nb/3.3)\% + (Ta/4)\% + (W/2)\% + (Cr/0.6)\% + (Mn/0.6)\% + (V/1.4)\% + (Fe/0.6)\% + (Co/0.9)\% + (Ni/0.8)\% \quad (6)$$

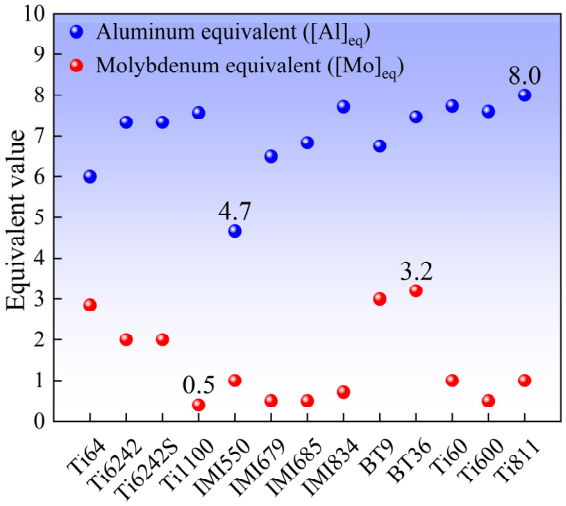

Figure 8. The values of $[Al]_{eq}$ and $[Mo]_{eq}$ of typical high-temperature titanium alloys.

Table 2. Alloy composition and prediction of UTS at 600 °C.

Alloy	Test Temperature (°C)	UTS (MPa)
Ti-7.2Al-1.8Mo-2.0Nb-0.4Si	600	649

3.4. Experimental Verification and Analysis

3.4.1. Validation of Prediction Results

The tensile stress–strain curves of the alloy at 600 °C and room temperature are shown in Figure 9, and the corresponding mechanical property values are listed in Table 3. As shown in Figure 9 and Table 3, the average values of UTS, YS, and EL of the alloy at 600 °C are 629 MPa (σ = 18.38 MPa), 478 MPa (σ = 14.19 MPa), and 15.5% (σ = 0.97%),

respectively. Equations (7) and (8) are used to calculate the experimental error and the prediction error of the XGBoost model. The results show that the experimental error is 2.9%, and the prediction error of the model for the UTS of the alloy at 600 °C is 3.2%. The analysis indicates that the model error and the experimental error are of the same order of magnitude, and both errors are relatively small. On one hand, the small experimental error fully demonstrates the reliability of the experimental data; on the other hand, the small prediction error reflects that the ML-based design method for high-temperature titanium alloys proposed in this study can accurately predict the mechanical properties, verifying the effectiveness of the design scheme. At room temperature, the alloy also exhibits relatively excellent performance, with average values of UTS, YS, and EL being 1035 MPa (σ = 14.76 MPa), 965 MPa (σ = 9.71 MPa), and 5.9% (σ = 0.21%), respectively.

$$\text{Error}_e = (\sigma/\text{UTS}_a) \times 100\% \quad (7)$$

$$\text{Error}_p = [(\text{UTS}_p - \text{UTS}_a)/\text{UTS}_a] \times 100\% \quad (8)$$

where Error_e is the experimental error of the model, σ is the standard deviation, UTS_a is the average value of the experimental UTS, Error_p is the prediction error of the model, and UTS_p is the predicted value of the UTS by the model.

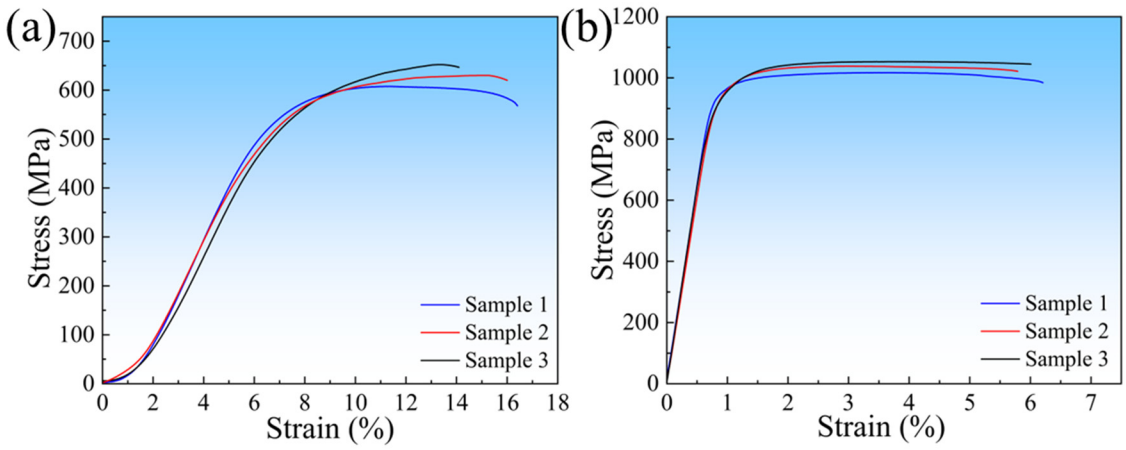

Figure 9. Tensile stress–strain curves of the as-cast alloy: (**a**) 600 °C; (**b**) room temperature.

Table 3. The mechanical properties of as-cast alloy at room temperature and 600 °C.

	600 °C			Room Temperature		
	UTS (MPa)	YS (MPa)	EL (%)	UTS (MPa)	YS (MPa)	EL (%)
Sample 1	607	465	16.3	1016	954	6.2
Sample 2	628	472	16.0	1037	968	5.7
Sample 3	652	498	14.1	1052	975	6.0
Average value	629	478	15.5	1035	965	5.9
Standard deviation σ	18.38	14.19	0.97	14.76	9.71	0.21

Table 4 presents the UTS data of the novel alloy and two typical as-cast high-temperature titanium alloys (Ti1100 and Ti6242S, both containing six elements). As shown in Table 4, the room-temperature UTS of the novel alloy is higher than that of the two typical high-temperature titanium alloys mentioned above, and its 600 °C high-temperature UTS is higher than that of Ti6242S, but slightly lower than that of Ti1100. The novel alloy has successfully reduced the number of alloying elements while ensuring high performance.

This fully demonstrates the excellent potential of ML-assisted design methods in predicting mechanical properties and optimizing compositions within the field of titanium alloy development, and opening up a new path for the efficient development of high-performance and low-cost titanium alloys.

Table 4. Comparison of UTS with two typical as-cast high-temperature titanium alloys.

Alloy	Test Temperature (°C)	UTS (MPa)
Ti-6Al-2.7Sn-4Zr-0.4Mo-0.45Si (Ti1100) [47]	20 600	930 650
Ti-6Al-2Sn-4Zr-2Mo-0.1Si (Ti6242S) [47]	20 600	1020 595
This work	20 600	1035 629

3.4.2. Feature Importance Analysis

In the field of ML research, the SHapley Additive exPlanations (SHAP) method transforms the "black-box" decisions of ML models into interpretable feature importance distributions by quantifying the contribution of each feature to the model's prediction results [48,49]. To understand the influence of input features (test temperature, alloy composition) on the UTS, this study used the SHAP method to analyze the feature importance of the XGBoost model. Figure 10 shows the SHAP analysis results of the XGBoost model, where red indicates higher feature values (composition content and test temperature), and blue indicates lower feature values. The horizontal axis represents the SHAP value, whose magnitude reflects the degree to which the feature affects the UTS. As shown in Figure 10, the test temperature is the most important feature affecting the UTS. SHAP values decrease with increasing test temperatures, indicating that the UTS of the alloy decreases with the increase in the test temperature. The feature importance ranking of alloy compositions is Al > Si > Mo > Nb, suggesting that Al is a key feature influencing the UTS. A higher Al content (red) corresponds to a more positive SHAP value, indicating that an increase in Al content is beneficial to improving the UTS. A similar trend was observed for Mo content. The SHAP values of Si and Nb do not show obvious patterns with their composition contents, and their contributions to the alloy's UTS are complex, suggesting that the effects of Si and Nb elements on the UTS of the alloy are largely influenced by other elements.

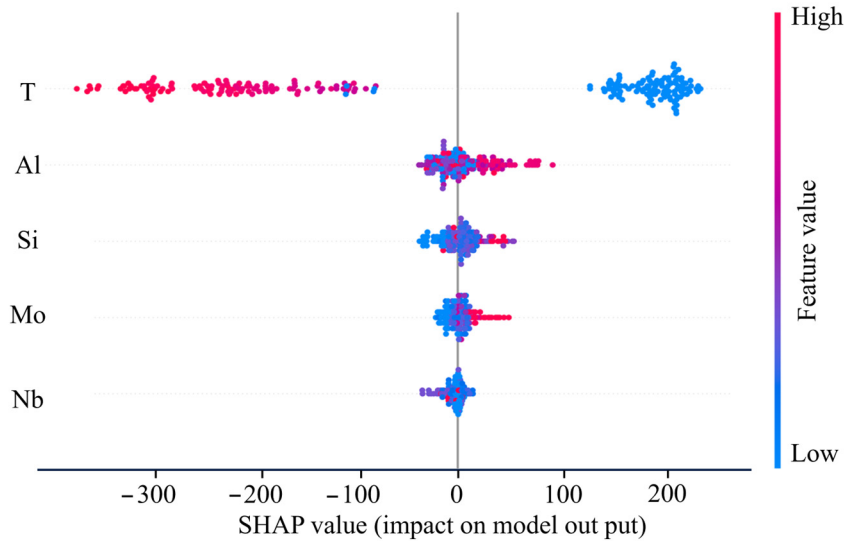

Figure 10. The SHAP analysis results of the XGBoost model.

3.4.3. Microstructure Characterization

Figure 11 shows the microstructure and the XRD diffraction pattern of the alloy. The microstructure of the alloy is a Widmanstätten structure, which is a typical microstructure of high-temperature titanium alloys. There is an obvious grain boundary α phase at the grain boundaries. At the same time, α/β bundles in different directions exist within the grains. This is because the α phase in the β grains needs to follow the 12 Burgers orientation relationships: $(0001)_\alpha//(110)_\beta$ and $<11\bar{2}0>_\alpha//<111>_\beta$ [50]. The quantitative statistical analysis of the microstructural characteristics was conducted using Image-Pro Plus software (V6.0, Media Cybernetics, Rockville, MD, USA) (average values calculated from measurements of 10 SEM images). The results show that the thickness of α-laths is 0.85 μm, the size of α/β colonies is 26.0 μm, and the volume fraction of the α phase is 75.4%. The analysis of the XRD diffraction pattern of the alloy shows that the phase of the alloy consists of an α-phase and a small amount of β-phase, classifying the designed alloy as an α+β type titanium alloy.

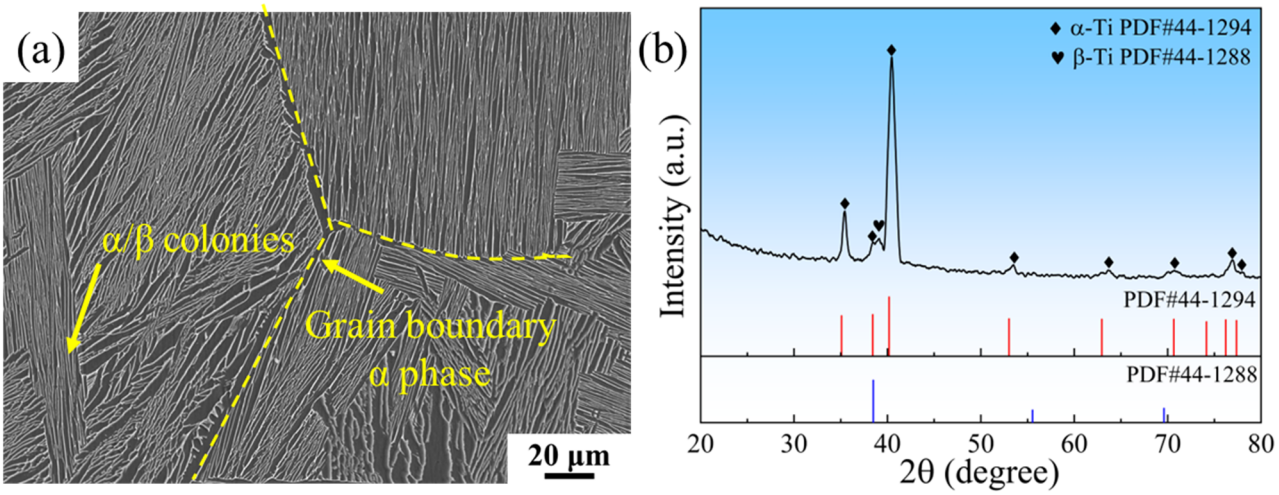

Figure 11. The microstructure and XRD diffraction pattern of the alloy: (**a**) SEM morphology; (**b**) XRD diffraction pattern.

3.4.4. Fracture Mechanism Analysis

Figure 12a shows the SEM morphology of the tensile fracture surface of the alloy at room temperature. Region A exhibits remarkable dimple characteristics, indicating the occurrence of ductile trans-granular fracture. Regions B and C are characterized by typical tearing ridge and river patterns, respectively, suggesting a brittle trans-granular fracture in these two regions. However, Region D exhibits a bright and reflective surface at the macroscopic level, and has the characteristic of tiny dimples at the microscopic level, indicating the occurrence of ductile intergranular fracture. Therefore, the fracture mode of the alloy at room temperature is a mixed fracture of brittleness and ductility. Figure 12b shows the SEM morphology of the tensile fracture surface of the alloy at 600 °C. Regions A and B exhibit the characteristics of dimples, pointing to ductile trans-granular fracture in these two regions. A crack is observed in Region C, while Region D shows the characteristics of cleavage plane, indicating that brittle trans-granular fracture occurred in these two regions. Therefore, the high-temperature fracture mode of the alloy is also a mixed fracture of brittleness and ductility.

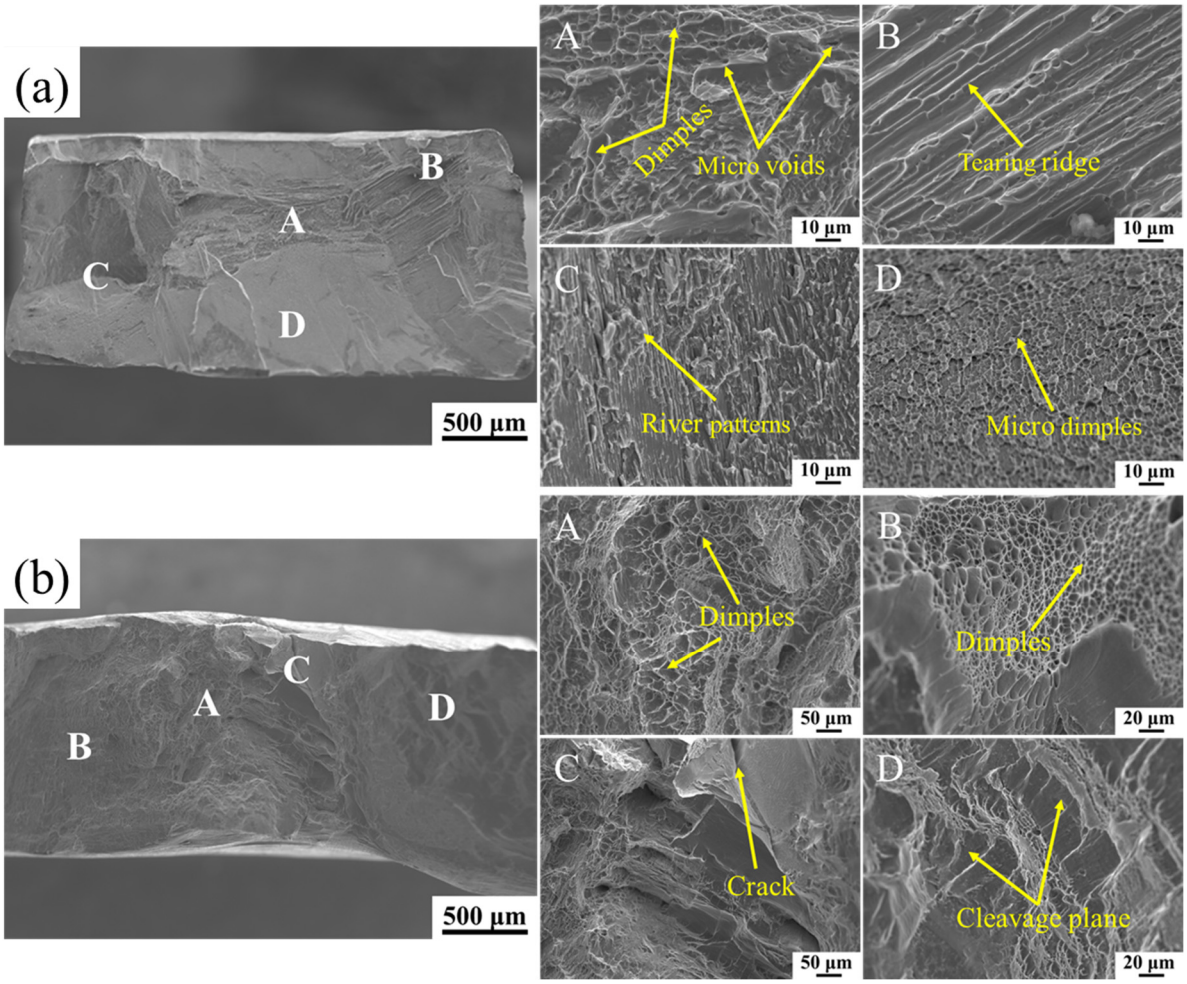

Figure 12. SEM morphology of tensile fracture surfaces of the alloy: (**a**) room temperature; (**b**) 600 °C.

The room-temperature fracture mechanism of the alloy is interpreted as follows: As the specimen deforms under the tensile load, the central position (Region A) elongates perpendicular to the tensile axis, thereby forming dimples and tiny pores. Driven by cumulative effect and stress concentration, these pores gradually propagate to the adjacent region, resulting in brittle trans-granular fracture in Region B and Region C. When the center of the specimen is completely separated, this failure will spread outward. Due to the change in the elastoplastic constraint, the maximum shear plane (at a 45° angle to the tensile axis) is likely to become the deformation plane, inducing ductile intergranular fracture at the grain boundaries (Region D). The high-temperature fracture mechanism of the alloy is interpreted as follows: The central position initially elongates perpendicular to the tensile axis. With the increase in the tensile temperature and the plastic strain reaching a critical level, obvious plastic deformation occurs, forming a large number of deep dimples in Region A. With the continued deformation, due to the material-softening effect caused by the high temperature, a ductile trans-granular fracture occurs in Region B (located on the side of the alloy), accompanied by pronounced shear fracture characteristics, resulting in a typical dimple morphology. The formation of the crack in region C arises from a large amount of deformation-induced significant inhomogeneous plastic deformation at high temperatures, triggering the initiation of the crack and the brittle trans-granular fracture in Region D.

4. Conclusions

This study successfully employed an ML method to design a novel high-temperature titanium alloy, Ti-7.2Al-1.8Mo-2.0Nb-0.4Si (wt.%), and validated its feasibility. The specific conclusions are as follows:

(1) Data augmentation significantly improved the performance of all four ML models. This indicated that data augmentation enabled the models to better learn the dataset information more effectively, thereby enhancing the generalization ability and prediction accuracy of the models.

(2) By training and testing the augmented dataset, the XGBoost model exhibited superior prediction performance with the R^2 of 0.94, RMSE of 53.31, and MAE of 42.93 for the test set.

(3) Combining the physical metallurgy model with the genetic algorithm improved the efficiency and accuracy of alloy design. The UTS, YS, and EL of the designed alloy at 600 °C were 629 MPa, 478 MPa, and 15.5%, respectively. The predicted UTS of the alloy at 600 °C by the XGBoost model was 649 MPa, with a minimal error of only 3.2%, which verified the efficacy of the machine learning method used in this study for designing novel high-temperature titanium alloys. At room temperature, the UTS, YS, and EL of the alloy were 1035 MPa, 965 MPa, and 5.9%, respectively.

(4) The microstructure analysis revealed that the designed high-temperature titanium alloy was of the α+β type, and its structure was a typical Widmanstätten structure. The fracture modes of the alloy at room temperature and high temperature were mixed fractures of brittleness and ductility.

This study has achieved remarkable results in the field of composition design for a novel quinary high-performance high-temperature titanium alloy through machine learning methods. Compared to as-cast Ti1100 and Ti6242S alloy (both containing six elements), the novel alloy has considerable high-temperature (600 °C) mechanical properties and fewer components. In the future, the continuous optimization of ML approaches will continue to drive the discovery and design of high-performance materials, providing a technical pathway that balances efficiency and reliability for the precise design of advanced structural materials.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ma18133099/s1, Table S1: Original data; Table S2: Augmented data.

Author Contributions: Conceptualization, B.F.; methodology, B.F.; validation, X.F., B.L. and T.D.; formal analysis: J.L.; investigation, X.F.; resources, B.L.; data curation, X.F. and T.D.; writing—original draft preparation, X.F. and B.F.; writing—review and editing, B.F. and B.L.; visualization, T.D.; supervision, T.D. and B.L.; project administration, J.L.; funding acquisition, B.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the S&T Program of Tianjin under grant No. 22YDTPJC00150.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in this article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Qu, B.L.; Zhang, C.J.; Lian, Q.H.; Deng, Y.L.; Zhang, S.Z.; Du, Z.X.; Han, J.C.; Wang, B.; Wang, T.; Zhang, X.Y. A novel near-α high temperature titanium alloy with trimodal microstructure and submicron-nanosilicides for superior mechanical properties at both room and elevated temperatures. *Compos. Part B* **2025**, *299*, 112411. [CrossRef]
2. Li, J.; Xu, Y.Q.; Xiao, W.L.; Ma, C.L.; Huang, X. Development of Ti-Al-Ta-Nb-(Re) near-α high temperature titanium alloy: Microstructure, thermal stability and mechanical properties. *J. Mater. Sci. Technol.* **2022**, *109*, 1–11. [CrossRef]
3. Li, T.Y.; Zhu, Z.H.; Zhang, S.; Min, X.H.; Dong, C. Effect of processing parameters on formability, microstructure, and microhardness of a novel laser additive manufactured Ti-6.38Al-3.87V-2.43Mo alloy. *China Foundry* **2022**, *19*, 158–168. [CrossRef]
4. Sun, P.; Su, B.L.; Tan, C.S.; Tang, B.; Li, J.H.; Yuan, R.H. The influence of heat treatment on the high temperature tensile properties of a near-α as-cast titanium alloy. *J. Alloys Compd.* **2025**, *1010*, 178148. [CrossRef]
5. Liu, T.Y.; Shi, K.; Zhao, J.; Liu, S.B.; Zhang, Y.W.; Liu, H.Y.; Chen, X.M.; Mei, X.M. Effect of hot isostatic pressing processing parameters on microstructure and properties of Ti60 high temperature titanium alloy. *China Foundry* **2023**, *20*, 49–56. [CrossRef]
6. Song, Y.J.; Fu, B.G.; Dong, T.S.; Li, G.L.; Wang, F.; Zhao, X.B.; Liu, J.H. Effect of Nb content on cyclic oxidation behavior of as-cast Ti-1100 alloys. *Materials* **2020**, *13*, 1082. [CrossRef]
7. Shen, J.F.; Fu, B.G.; Wang, Y.F.; Dong, T.S.; Li, J.K.; Li, G.L.; Liu, J.H. Effect of Y content on precipitation behavior, oxidation and mechanical properties of as-cast high-temperature titanium alloys. *Materials* **2023**, *16*, 4784. [CrossRef]
8. Ma, X.Z.; Chen, Z.Y.; Xiang, Z.L.; Zhang, S.W.; Liu, Y.Y.; Ding, X.X. Influence of quenching plus aging on microstructures and mechanical properties evolutions for Ti-Al-Sn-Zr-Mo-Si-Nb-Ta-Er-C near-α high temperature titanium alloys. *Mater. Today Commun.* **2024**, *38*, 108457. [CrossRef]
9. Yan, W.G.; Wang, H.M.; Tang, H.B.; Cheng, X.; Zhu, Y.Y. Effect of Nd addition on microstructure and tensile properties of laser additive manufactured TC11 titanium alloy. *Trans. Nonferrous Met. Soc. China* **2022**, *32*, 1501–1512. [CrossRef]
10. Li, C.Z.; Fu, B.G.; Dong, T.S.; Liu, J.H.; Song, Y.J.; Zhao, X.B.; Li, G.L. Microstructure and dry sliding wear behavior of as-cast TiC_p/Ti-1100-0.5Nb titanium matrix composite at elevated temperatures. *China Foundry* **2020**, *17*, 455–463. [CrossRef]
11. Bodunrin, M.O.; Chown, L.H.; Omotoyinbo, J.A. Development of low-cost titanium alloys: A chronicle of challenges and opportunities. *Mater. Today Proc.* **2021**, *38*, 564–569. [CrossRef]
12. Sengupta, B.; Shekhar, S.; Kulkarni, K.N. A novel ultra-high strength and low-cost as-cast titanium alloy. *Mater. Sci. Eng. A* **2017**, *696*, 478–481. [CrossRef]
13. Huang, S.X.; Zhao, Q.Y.; Yang, Z.; Lin, C.; Zhao, Y.Q.; Yu, J.S. Strengthening effects of Al element on strength and impact toughness in titanium alloy. *J. Mater. Res. Technol.* **2023**, *26*, 504–516. [CrossRef]
14. Lu, J.W.; Ge, P.; Zhao, Y.Q. Recent Development of effect mechanism of alloying elements in titanium alloy design. *Rare Metal Mat. Eng.* **2014**, *43*, 775–779. [CrossRef]
15. Kotov, A.D.; Postnikova, M.N.; Mosleh, A.O.; Mikhaylovskaya, A.V. Effect of Mo content on the microstructure, superplastic behavior, and mechanical properties of Ni and Fe-modified titanium alloys. *Mater. Sci. Eng. A* **2023**, *877*, 145166. [CrossRef]
16. Huang, S.X.; Zhao, Q.Y.; Zhao, Y.Q.; Lin, C.; Wu, C.; Jia, W.J.; Mao, C.L.; Ji, V. Toughening effects of Mo and Nb addition on impact toughness and crack resistance of titanium alloys. *J. Mater. Sci. Technol.* **2021**, *79*, 147–164. [CrossRef]
17. Cui, Y.J.; Aoyagi, K.; Koizumi, Y.; Yang, C.; Bian, H.K.; Hayasaka, Y.; Fujieda, T.; Chiba, A. Effect of niobium addition on tensile properties and oxidation resistance of a titanium-based alloy. *Corros. Sci.* **2021**, *180*, 109198. [CrossRef]
18. Marker, C.; Shang, S.L.; Zhao, J.C.; Liu, Z.K. Thermodynamic description of the Ti-Mo-Nb-Ta-Zr system and its implications for phase stability of Ti bio-implant materials. *Calphad-Comput. Coupling Ph. Diagrams Thermochem.* **2018**, *61*, 72–84. [CrossRef]
19. Fu, B.G.; Wang, H.W.; Zou, C.M.; Wei, Z.J. The effects of Nb content on microstructure and fracture behavior of near α titanium alloys. *Mater. Des.* **2015**, *66*, 367–373. [CrossRef]
20. Liu, Z.M.; Zhu, P.L.; Mao, X.N.; Zhao, Y.Q.; Xin, S.W. Effect of Si addition on the microstructure and creep properties of the forged titanium alloy. *Mater. Chem. Phys.* **2024**, *317*, 129212. [CrossRef]
21. Guo, R.; Liu, B.; Xu, R.J.; Cao, Y.K.; Qiu, J.W.; Chen, F.; Yan, Z.Q.; Liu, Y. Microstructure and mechanical properties of powder metallurgy high temperature titanium alloy with high Si content. *Mater. Sci. Eng. A* **2020**, *777*, 138993. [CrossRef]
22. Fu, B.G.; Wang, H.W.; Zou, C.M.; Wei, Z.J. The influence of Zr content on microstructure and precipitation of silicide in as-cast near α titanium alloys. *Mater. Charact.* **2015**, *99*, 17–24. [CrossRef]
23. Zhuo, L.C.; Ji, K.L.; Lu, J.W.; Sun, J.C.; Huo, W.T.; Shao, H.; Chen, B.Q.; Zhao, Y.Q. Microstructure characterization and tensile performance of a high-strength titanium alloy with in-situ precipitates of Ti_5Si_3. *J. Alloys Compd.* **2023**, *968*, 171867. [CrossRef]
24. Kishida, K.; Fujiwara, M.; Adachi, H.; Tanaka, K.; Inui, H. Plastic deformation of single crystals of Ti_5Si_3 with the hexagonal $D8_8$ structure. *Acta Mater.* **2010**, *58*, 846–857. [CrossRef]
25. Gao, X.Y.; Wang, H.Y.; Tan, H.J.; Xing, L.; Hu, Z.Y. Data-driven machine learning for alloy research: Recent applications and prospects. *Mater. Today Commun.* **2023**, *36*, 106697. [CrossRef]
26. Liu, X.J.; Xu, P.C.; Zhao, J.J.; Lu, W.C.; Li, M.J.; Wang, G. Material machine learning for alloys: Applications, challenges and perspectives. *J. Alloys Compd.* **2022**, *921*, 165984. [CrossRef]

27. Hu, M.W.; Tan, Q.Y.; Knibbe, R.; Xu, M.; Jiang, B.; Wang, S.; Li, X.; Zhang, M.X. Recent applications of machine learning in alloy design: A review. *Mater. Sci. Eng. R-Rep.* **2023**, *155*, 100746. [CrossRef]
28. Liu, Z.D.; Zhang, H.Y.; Zhang, S.; Cheng, J.; He, Y.X.; Zhou, G.; Liu, J.W.; Song, S.P.; Chen, J.J. A machine learning method approach for designing novel high strength and plasticity metastable β titanium alloys. *Prog. Nat. Sci. Mater. Int.* **2025**, *35*, 156–165. [CrossRef]
29. Wu, C.T.; Chang, H.T.; Wu, C.Y.; Chen, S.W.; Huang, S.Y.; Huang, M.X.; Pan, Y.T.; Bradbury, P.; Chou, J.; Yen, H.W. Machine learning recommends affordable new Ti alloy with bone-like modulus. *Mater. Today.* **2020**, *34*, 41–50. [CrossRef]
30. Oh, J.K.; Narayana, P.L.; Hong, J.K.; Yeom, J.T.; Reddy, N.S.; Kang, N.; Park, C.H. Property optimization of TRIP Ti alloys based on artificial neural network. *J. Alloys Compd.* **2021**, *884*, 161029. [CrossRef]
31. Nawa, K.; Hagiwara, K.; Nakamura, K. Prediction-accuracy improvement of neural network to ferromagnetic multilayers by Gaussian data augmentation and ensemble learning. *Comput. Mater. Sci.* **2023**, *219*, 112032. [CrossRef]
32. Ye, Y.C.; Li, Y.H.; Ouyang, R.L.; Zhang, Z.R.; Tang, Y.; Bai, S.X. Improving machine learning based phase and hardness prediction of high-entropy alloys by using Gaussian noise augmented data. *Comput. Mater. Sci.* **2023**, *223*, 112140. [CrossRef]
33. Vela, B.; Khatamsaz, D.; Acemi, C.; Karaman, I.; Arróyave, R. Data-augmented modeling for yield strength of refractory high entropy alloys: A Bayesian approach. *Acta Mater.* **2023**, *261*, 119351. [CrossRef]
34. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Trans. Evol. Comput.* **1997**, *1*, 67–82. [CrossRef]
35. Chai, C.R.; Wang, Y.; Zhao, S.; Zhang, Y.X.; Fang, F.; Peng, L.; Zhang, X.M. Machine learning-assisted design of low elastic modulus β-type medical titanium alloys and experimental validation. *Comput. Mater. Sci.* **2024**, *238*, 112902. [CrossRef]
36. *GB/T 228.2-2015*; Metallic Materials—Tensile Testing—Part 2: Method of Test at Elevated Temperature. The National Standards of the People's Republic of China: Beijing, China, 2015.
37. *ASTM E8/E8M-16*; Standard Test Methods for Tension Testing of Metallic Materials. ASTM International: West Conshohocken, PA, USA, 2020.
38. Zhang, Y.; Li, D.S.; Li, X.Q.; Liu, X.C.; Zhao, S.T.; Li, Y. Creep deformation and strength evolution mechanisms of a Ti-6Al-4V alloy during stress relaxation at elevated temperatures from elastic to plastic loading. *J. Mater. Sci. Technol.* **2022**, *126*, 93–105. [CrossRef]
39. Liu, Z.Y.; Wang, T.Y.; Jin, L.; Zeng, J.; Dong, S.; Wang, F.H.; Wang, F.L.; Dong, J. Towards high stiffness and ductility-The Mg-Al-Y alloy design through machine learning. *J. Mater. Sci. Technol.* **2025**, *221*, 194–203. [CrossRef]
40. Wang, Z.; Zhang, H.Y.; Wang, S.Y.; Cheng, J.; Wang, C.; Zhou, G.; Zhang, S.Q.; Chen, L.J. A machine learning method based on TPE-XGBoost model for TRIP/TWIP near-β titanium alloy design. *Mater. Today Commun.* **2024**, *41*, 110309. [CrossRef]
41. Wang, H.N.; Sun, Q.Z.; Wu, J.; Zhang, X.X.; Liu, W.P.; Peng, T.; Tang, R.Z. Self-training-based approach with improved XGBoost for aluminum alloy casting quality prediction. *Robot. Comput. Integr. Manuf.* **2025**, *92*, 102890. [CrossRef]
42. Gysler, A.; Weissmann, S. Effect of order in Ti_3Al particles and of temperature on the deformation behavior of age-hardened Ti-Al alloys. *Mater. Sci. Eng.* **1977**, *27*, 181–193. [CrossRef]
43. Li, J.; Cai, J.M.; Xu, Y.Q.; Xiao, W.L.; Huang, X.; Ma, C.L. Influences of thermal exposure on the microstructural evolution and subsequent mechanical properties of a near-α high temperature titanium alloy. *Mater. Sci. Eng. A* **2020**, *774*, 138934. [CrossRef]
44. Cao, S.; Zhang, S.Z.; Liu, J.R.; Li, S.J.; Sun, T.; Li, J.P.; Gao, Y.; Yang, R.; Hu, Q.M. Interaction between Al and other alloying atoms in α-Ti for designing high temperature titanium alloy. *Comput. Mater. Sci.* **2021**, *197*, 110620. [CrossRef]
45. Li, Y.L.; Fang, H.Z.; Chen, R.R.; Sun, S.C.; Xue, X.; Guo, J.J. Optimization of (α+β) microstructure and trade-off between strength and toughness: Based on $Mo_{[eq]}$ and d electron theory in β-Ti alloy. *Mater. Des.* **2023**, *231*, 112022. [CrossRef]
46. Zhao, E.T.; Sun, S.C.; Zhang, Y. Recent advances in silicon containing high temperature titanium alloys. *J. Mater. Res. Technol.* **2021**, *14*, 3029–3042. [CrossRef]
47. Bania, P.J. Next generation titanium alloys for elevated temperature service. *ISIJ Int.* **1991**, *31*, 840–847. [CrossRef]
48. Huang, J.X.; Ando, D.; Sutou, Y. Heat-resistant aluminum alloy design using explainable machine learning. *Mater. Des.* **2024**, *243*, 113057. [CrossRef]
49. Dong, Z.Q.; Zhou, C.; Huang, Q.L.; Mou, Z.K.; Li, M.; Zhou, H.Y.; Zheng, W.Y. Design of high temperature oxidation-resistant high-entropy alloys via machine learning and natural mixing process. *Corros. Sci.* **2025**, *225*, 113047. [CrossRef]
50. Zhang, J.Y.; Zhang, Y.S.; Mompiou, F.; Zhang, W.Z. Structures and migrations of interfaces between β precipitates and α′ matrix in a Ti-2.6 wt% Mo alloy. *Acta Mater.* **2024**, *281*, 120429. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.