Acta Materialia 195 (2020) 454–467

Contents lists available at ScienceDirect

Acta Materialia

journal homepage: www.elsevier.com/locate/actamat

Full length article

Predicting creep rupture life of Ni-based single crystal superalloys using divide-and-conquer approach based machine learning

Yue Liu[a], Junming Wu[a], Zhichao Wang[b], Xiao-Gang Lu[b,c], Maxim Avdeev[d,e], Siqi Shi[b,c,*], Chongyu Wang[f], Tao Yu[g]

[a] School of Computer Engineering and Science, Shanghai Engineering Research Center of Intelligent Computing System, Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai 200444, China
[b] State Key Laboratory of Advanced Special Steel, Shanghai Key Laboratory of Advanced Ferrometallurgy, School of Materials Science and Engineering, Shanghai University, Shanghai 200444, China
[c] Materials Genome Institute, Shanghai University, Shanghai 200444, China
[d] Australian Nuclear Science and Technology Organisation, Locked Bag 2001, Kirrawee DC, NSW 2232, Australia
[e] School of Chemistry, The University of Sydney, Sydney 2006, Australia
[f] Department of Physics, Tsinghua University, Beijing 100084, China
[g] Institute of Functional Materials, General Institute of Iron and Steel Research, Beijing 100084, China

ARTICLE INFO

Article History:
Received 3 November 2019
Revised 4 April 2020
Accepted 1 May 2020
Available online 17 May 2020

Keywords:
Ni-based single crystal superalloy
Creep property prediction
Machine learning
Alloy design

ABSTRACT

Creep rupture life is a key material parameter for service life and mechanical properties of Ni-based single crystal superalloy materials. Therefore, it is of much practical significance to accurately and efficiently predict creep life. Here, we develop a divide-and-conquer self-adaptive (DCSA) learning method incorporating multiple material descriptors for rational and accelerated prediction of the creep rupture life. We characterize a high-quality creep dataset of 266 alloy samples with such features as alloy composition, test temperature, test stress, and heat treatment process. In addition, five microstructural parameters related to creep process, including stacking fault energy, lattice parameter, mole fraction of the γ' phase, diffusion coefficient and shear modulus, are calculated and introduced by the CALPHAD (CALculation of PHAse Diagrams) method and basic materials structure-property relationships, that enables us to reveal the effect of microstructure on creep properties. The machine learning explorations conducted on the creep dataset demonstrate the potential of the approach to achieve higher prediction accuracy with *RMSE, MAPE* and R^2 of 0.3839, 0.0003 and 0.9176 than five alternative state-of-the-art machine learning models. On the newly collected 8 alloy samples, the error between the predicted creep life value and the experimental measured value is within the acceptable range (6.4486 h–40.7159 h), further confirming the validity of our DCSA model. Essentially, our method can establish accurate structure-property relationship mapping for the creep rupture life in a faster and cheaper manner than experiments and is expected to serve for inverse design of alloys.

1. Introduction

Ni-based single crystal superalloys have become the most widely used material for gas-turbine blades of aero engines owing to its remarkable mechanical properties [1]. However, the service life of alloy products is limited by creep properties of materials, as creep deformation and rupture of materials under stress at high temperature lead to failure of products [2]. Therefore, creep rupture life is an important physical parameter in the design of new alloys. Although the creep rupture life of superalloy can be determined by experimental methods, the measurements are costly and time-consuming due to long-term creep testing and expensive alloy manufacturing [3], in particular when it involves precious metal elements such as Re, Ru, Ir, etc. [4].

Several theoretical methods have been presented to accelerate prediction of creep rupture life of alloys. They can be grouped into two categories including time-temperature parameter (TTP) methods and creep constitutive models (CCM). As an example of using TTP method, Larson–Miller [5] estimates the long-term creep rupture life of alloys at low temperatures based on the short-term creep test data. Similarly, based on the data of steady-state creep rate and rupture time under the same temperature and different stress conditions, isothermal extrapolation method [6] establishes empirical relationship between stress and rupture time and extrapolates the creep behavior to long-term conditions. However, these TTP methods, being based on empirical analysis of experimental data, lack a rigorous theoretical basis. Moreover, they fail to fully consider the

* Corresponding author.
E-mail address: sqshi@shu.edu.cn (S. Shi).

https://doi.org/10.1016/j.actamat.2020.05.001
1359-6454/

microstructure evolution information in the actual creep process. With the generalization of the theory of crystal plasticity and continuum mechanics, creep constitutive models for predicting creep rupture life have been developing rapidly in the recent years. MacLachlan and Knowles [7] simulated and predicted the stress fracture properties of four different types of superalloys using a modified continuum damage model (CDM) coupled with the descriptors related to creep deformation and microstructure changes. Feng et al. [8] developed an anisotropic continuum damage model for explaining creep damage, which can not only measure the damage degree, but also consider the anisotropy of the damage. Prasad et al. [9] modeled the creep behavior under different loads from the damage caused by dislocation motion and interfacial dislocation network. Vladimirov et al. [10] considered the dislocation movement characteristics at each stage of creep of single-crystal alloy, and established a model of the creep behavior of single-crystal alloy. Fedelich et al. [11] introduced microstructure parameter to take into account the changes caused by γ' phase rafting, and their effects on the creep of the alloy. Considering the characteristics of microstructure evolution and dislocation motion during creep process, it not only gives the model a more explicit physical meaning, but also more accurately describes the creep behavior and improves the prediction accuracy of creep life. In summary, the modern theoretical methods can accurately describe the creep process and be used to successfully predict creep rupture life of the alloy in a shorter time compared to experimental measurements. In order to further accelerate modelling of creep properties, Kim et al. [12] described high-throughput calculations combining theoretical models with the CALPHAD (CALculation of PHAse Diagrams) method. However, the approach is limited to intermediate temperature range and is not suitable for studying creep at high temperatures.

Recently, data-driven machine learning methods have been successfully employed for materials properties prediction [13–15], new materials discovery [16] or other purposes [17], and have proved their superiority in time efficiency and prediction performance [18]. Given that machine learning is the method of choice to reveal complex relationships between material descriptors and properties of interest, it is a very promising tool to be utilized for creep property prediction of Ni-based single crystal superalloy. Until now, some efforts have been made to the applications of statistical models in the prediction of creep rupture life of superalloy. Venkatesh and Rack [19] employed a BPNN (back-propagation neural network) to predict the creep rupture life of INCONEL 690 Ni-based single crystal superalloy at 1000 °C and 1100 °C. When the acceptable lifetime error range is set between −2 h and 2 h, the prediction accuracies of the models are 100% and 90%, respectively. Yoo et al. [20] utilized Bayesian neural network with Markov chain Monte Carlo methods to investigate the mechanism of creep properties of single crystal superalloy using relevant material descriptors including chemical composition, test stress and test temperature, and predictive accuracy of the model reached 93.2%.

Therefore, the feasibility of these data-driven models in estimating creep rupture life of Ni-based single crystal superalloy has been successfully demonstrated. However, previous efforts have largely focused only on basic physical factors such as chemical composition, test conditions and heat treatment parameters and did not take into account the factors related to microstructural evolution information, such as lattice parameter [21], mole fraction of γ' phase [22], shear modulus [23], stacking fault energy [22] and diffusion coefficient [12], which are known to be highly relevant to the alloy creep behavior. Furthermore, it is widely accepted that the mechanical and creep properties of different alloys at high temperatures correspond to the various different deformation mechanisms [22], thus making it difficult to capture the relationships among the relevant material descriptors and creep properties in these alloys with mixed creep mechanisms through a single learning model. And no statistical model in the prediction of creep rupture life of superalloy takes into consideration these differences of creep mechanisms of alloys.

In order to construct a general machine learning model to predict creep behavior, in this study we propose a novel divide-and-conquer self-adaptive (DCSA) learning method. In addition to the basic descriptors, such as chemical composition, test conditions and heat treatment process parameters, five microstructural factors (lattice parameter, mole fraction of γ' phase, shear modulus, stacking fault energy and diffusion coefficient) that can reflect the creep deformation mechanism were calculated and introduced using the CALPHAD method and material structure-property relationships. Moreover, instead of learning complex creep mechanisms through a single regression method, the DCSA method adopts the idea of “divide and conquer” methodology. Firstly, the DCSA automatically classifies alloy samples with various creep mechanisms into several clusters based on comprehensive factors (chemical composition, test conditions, heat treatment, and five calculated microstructural factors) through K-Means clustering. Secondly, based on the designed fitness function, the DCSA self-adaptively chooses the optimal machine learning model from the five candidates, i.e. Random Forest Regression (RF), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Lasso Regression (LR) and Ridge Regression (RR), to reveal the differences in creep mechanisms of alloy samples in different clusters. Finally, we illustrate the process and its performance using the dataset collected from the journal literature and patents and show that compared to the five state-of-the-art regression models Random Forest regression (RF), Support Vector Regression (SVR), Gaussian Process Regression (GPR), Lasso Regression (LR) and Ridge Regression (RR)), the DCSA achieves better prediction performance with *RMSE, MAPE,* and R^2 of 0.3839, 0.0003 and 0.9176, respectively. Further, in order to demonstrate the generalization ability of our model on unseen samples, we conduct experiments on 8 newly collected alloy samples, and the results show that the error between the predicted values of the model and the experimental measured values is within the acceptable range (6.4486 h−40.7159 h).

The remainder of the paper is structured as follows. Section 2 introduces the working steps of using the Thermo-Calc method and basic structure-property relations to calculate five microstructural descriptors related to the creep process. Section 3 describes the details of the proposed method. Section 4 demonstrates the performance of the method through plenty experiments on the collected creep dataset. Finally, the conclusions of this study are given in Section 5.

2. Microstructural descriptor calculations based on the CALPHAD method and phenomenological models

Ni-based single crystal superalloys are characterized by high content of cubic γ' precipitate ($L1_2$, ordered face-centered cubic Ni_3Al) embedded in the γ matrix (disordered fcc, Ni-based solid solution) [24]. Creep in γ/γ' microstructure involves a number of elementary processes including diffusion, dislocation glide, dislocation climb, pairwise cutting of γ'-particles and coarsening of γ'-particles [25–27]. To examine the effects of microstructural factors such as dislocation motion on creep behavior of superalloys, Zhang et al. [25] monitored the evolution of their microstructures by transmission electron microscopy (TEM). The morphologies of dislocations in the superalloys TMS-75 (+Ru) [28] and TMS-138 [29] after 2 h creep were studied. At this early stage of creep, the dislocation morphologies in TMS-138 are considerably different from those in TMS-75 (+Ru). As can be seen in Fig. 1(a) (see Fig. 1 in Ref. [25]), dislocations are observed in most of the area of this specimen and creep dislocations are gliding and climbing into the previously dislocation-free areas. In Fig. 1(b) (see Fig. 1 in Ref. [25]), the characteristics of the dislocations, such as the considerable length of the dislocation lines

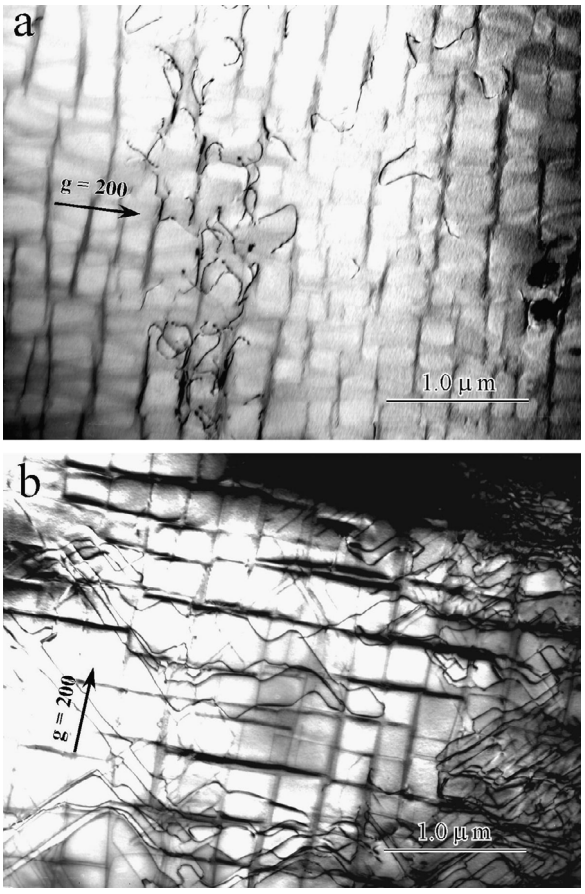

Fig. 1. Microstructural evolution during creep (1100 °C/137 MPa); t = 2 h [25]. (a) TMS-75(+Ru). Dislocations are moving by a process of gliding and climbing. (b) TMS-138. Dislocations are moving by cross-slip.

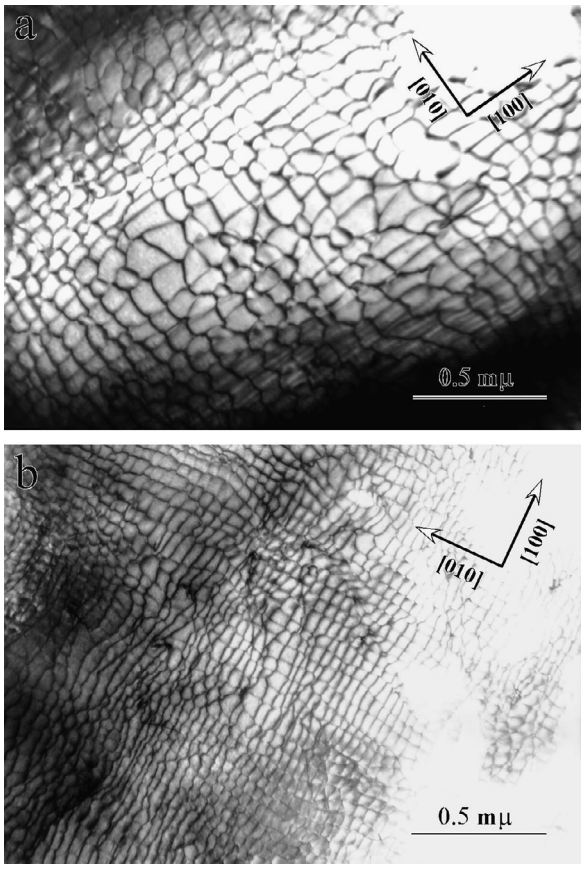

Fig. 2. Interfacial dislocation networks after creep rupture (1100 °C/137 MPa) [25]. (a) TMS-75(+Ru); (b) TMS-138.

which are parallel to the (001) plane, narrow and serrated, indicate the dislocations move by cross-slip. The morphologies of the γ/γ' interfacial dislocation networks in these two superalloys after creep rupture are schematically illustrated in Fig. 2 (see Fig. 2 in Ref. [25]). It is clearly seen that the dislocation network is denser in TMS-138 (Fig. 2(b)) than in TMS-75(+Ru) (Fig. 2(a)). The denser the dislocation network, the smaller is the minimum creep rate. This also indicates that TMS-138 alloy is superior to TMS-75(+Ru) in terms of creep properties. It is clear that microstructural information associated with dislocation morphology is important for understanding and predicting creep behavior. However, it is not experimentally available or difficult to quantify in many alloy datasets. Achieving accurate and quantitative predictions of microstructure evolution remains an open challenge, especially in the complex multi-component/multi-phase alloy systems. Herein, inspired by the creep model proposed by Kim et al. [12], we choose five critical microstructural factors, including the mole fraction of γ' precipitates, diffusion coefficient, shear modulus, lattice parameter, and stacking fault energy to indirectly quantify complex microstructural evolution behavior during creep process. The calculation of the five factors is described in detail in the following section.

The Thermo-Calc software [30], implementing the CALPHAD models and integrated with assessed thermodynamic databases, is a very powerful tool to calculate various thermodynamic properties of different materials, even for complex multi-component alloy systems containing more than eight alloying elements, such as Ni-based single crystal superalloy [31]. All of the CALPHAD thermodynamic calculations are assumed to be exact, although it should be noted that the errors in the experimental and *ab initio* calculation data or its sparsity lead to uncertainty of the determined interaction parameters stored in the thermodynamic databases and in turn to uncertainty of CALPHAD calculations [32–34].

In this study, for given element compositions and equilibrium temperature or test temperature, the phase amount (mole fraction), phase composition and site fraction of elements of alloys can be firstly obtained by Thermo-Calc. Then combined with phenomenological material models, which are discussed below, the values of the five factors (stacking fault energy, lattice parameter, mole fraction of the γ' phase, diffusion coefficient and shear modulus) are calculated.

The computational procedure is illustrated in Fig. 3 in detail. Firstly, the original creep dataset covering the chemical composition (Ni, Re, Co, Cr, Al, Ta, Ti, W, Mo, C, B, Y, Hf and Nb), solution treatment time and temperature, aging treatment time and temperature and test conditions is collected from the publications (as described in detail in Section 4.1). Secondly, several micro-parameters such as the phase composition, site fraction of elements and mole fraction of γ' phase$X^{\gamma'}$, related to the phase composition of alloys, are calculated by the Thermo-Calc software. Further, diffusion coefficient D_L, shear modulus G, stacking fault energy Γ and lattice parameter L, closely related to the creep process, are calculated via phenomenological material models. All the calculation methods or expressions about five microstructural factors are given in the following parts.

The γ' phase is the main strengthening phase in Ni-based single crystal superalloy, and its quantity, size, morphology and distribution are closely related to the mechanical and creep properties of the alloy [35]. The Thermo-Calc thermodynamic calculation method is used to calculate the mole fractions of the γ and γ' phase and further determines the weighted average lattice constant of the alloy as given in Eq. (1).

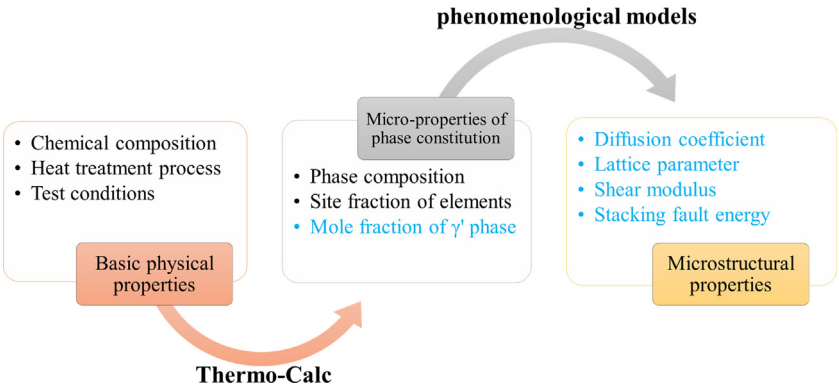

Fig. 3. The computational procedure of microstructural descriptors.

$$L=\left(1-X^{\gamma'}\right)L^{\gamma}+X^{\gamma'}L^{\gamma'} \tag{1}$$

where $X^{\gamma'}$ represents the mole fraction of γ' phase, obtained from the Thermo-Calc calculations. L^{γ} and $L^{\gamma'}$ represent the lattice parameters of γ and γ' phase respectively, which are given in Eqs. (2) and (3).

$$L^{\gamma}=L_{\mathrm{Ni}}*\left(1+\Delta_L^{\gamma}(x)\right) \tag{2}$$

$$L^{\gamma'}=L_{\mathrm{Ni_3Al}}*\left(1+\Delta_L^{\gamma'}(x)\right) \tag{3}$$

$$\Delta_L^{\gamma}(x)=\sum_i x_i\delta L_i \tag{4}$$

$$\Delta_L^{\gamma'}(x)=\sum_i\left(0.75\times y_i^{Nisite}\delta L_i^{Nisite}+0.25\times y_i^{Alsite}\delta L_i^{Alsite}\right) \tag{5}$$

where L_{Ni} and $L_{\mathrm{Ni_3Al}}$ are the lattice parameters for pure Ni and perfect Ni_3Al respectively. x_i is the mole fraction of solute element i in γ phase. δL_i represents the amount of change in lattice parameter of pure Ni with an addition of alloying element i. y_i represents site fraction in the Ni or Al site in γ' phase, and δL_i^{Nisite} and δL_i^{Alsite} represent the amount of change in lattice parameter of Ni_3Al when an alloying element i is added to the Ni and Al site in γ' phase. $\Delta_L^{\gamma}(x)$ and $\Delta_L^{\gamma'}(x)$ are the fractional changes of lattice parameter in γ and γ' phase due to alloying respectively, as shown in Eqs. (4) and (5). The model of the fractional changes is adopted from the work by Kim et al. [12] and used here to account for the alloying modified lattice parameter, as well as the following shear modulus and stacking fault energy.

Kassner and Perez-Prado [23] analyzed the flow stress and elastic modulus data of different metals and found that there is a linear relationship between them, indicating that the elastic modulus has an effect on the creep of superalloys. It is worth pointing out that some studies use Young's modulus E when describing creep problems, while others use shear modulus G. The difference between the two lies in a factor of $2^*(1+\nu)$, where ν is Poisson's ratio. Creep experiments are usually performed under unidirectional tensile or compression conditions, so it is more convenient to use E when processing macro experimental data. However, this paper deals with creep problems from a theoretical and microscopic perspective. Therefore, the shear modulus G is used, which is more reasonable because plastic deformation is achieved by shear strain on the sliding plane. The shear modulus G can be estimated from the corresponding elastic constants of γ or γ' phases and the Voigt model [36] is chosen for this purpose in this paper. Based on the widely accepted concept of "mixed rule", the shear modulus G of an alloy is defined in Eq. (6).

$$G=\left(1-X^{\gamma'}\right)*G^{\gamma}+X^{\gamma'}*G^{\gamma'} \tag{6}$$

where G^{γ} and $G^{\gamma'}$ represent shear modulus of γ and γ' phases, respectively, which are given in Eqs. (7) and (8).

$$G^{\gamma}=G_{\mathrm{Ni}}*\left(1+\Delta_G^{\gamma}(x)\right) \tag{7}$$

$$G^{\gamma'}=G_{\mathrm{Ni_3Al}}*\left(1+\Delta_G^{\gamma'}(x)\right) \tag{8}$$

$$\Delta_G^{\gamma}(x)=\sum_i x_i\delta G_i \tag{9}$$

$$\Delta_G^{\gamma'}(x)=\sum_i\left(0.75\times y_i^{Nisite}\delta G_i^{Nisite}+0.25\times y_i^{Alsite}\delta G_i^{Alsite}\right) \tag{10}$$

where G_{Ni} and $G_{\mathrm{Ni_3Al}}$ represent shear modulus for pure Ni and perfect Ni_3Al, respectively. x_i is the mole fraction of solute element i in γ phase. $\Delta_G^{\gamma}(x)$ and $\Delta_G^{\gamma'}(x)$ denote the fractional changes of shear modulus in γ and γ' phases due to alloying respectively, as shown in Eqs. (9) and (10). The meaning of δG_i, δG_i^{Nisite} and δG_i^{Alsite} are similar to δL_i, δL_i^{Nisite} and δL_i^{Alsite} mentioned already.

Guo and Saunders [37] found a potential correlation between the stacking fault energy and creep of FCC metals through experiments. Although the mechanism by which the fault energy affects the creep rate controlled by dislocation climb is unclear, there are two possible explanations: the energy affects the dislocation climb rate or affects the creep dislocation substructure. In the creep process controlled by cross-slip dislocation, alloys with low fault energy are less likely to have cross-slip dislocations, resulting in better creep properties of the superalloys. The stacking fault energy is given in Eq. (11) [12].

$$\Gamma=\Gamma_{\mathrm{Ni}}*\left(1+\Delta_\Gamma^{\gamma}(x)\right) \tag{11}$$

$$\Delta_\Gamma^{\gamma}(x)=\sum_i x_i\delta\Gamma_i \tag{12}$$

where Γ_{Ni} is the stacking fault energy of pure Ni, which is obtained by fitting the results of first-principles [38,39] and $\Delta_\Gamma^{\gamma}(x)$ represents the fractional change of stacking fault energy in γ phase due to alloying, as shown in Eq. (11). The values of $\delta\Gamma_i$ can be obtained by fitting the experimental data, and their fitting values are listed in Ref. [12].

The diffusion coefficient is used to describe the diffusion behavior of the alloy. The atomic diffusion mechanism is one of the important causes of the creep deformation of alloys. During the creep process, there is also the directional movement of elements in the alloy. In this study, the activation energy for creep of metal is considered to be roughly equal to the activation energy for self-diffusion of the metal. In previous work, Saunders et al. [40] and Zhu et al. [41] took the lattice diffusion coefficient of γ phase as the effective diffusion coefficient in their high temperature creep models, although

obviously this is a simplified way to calculate the diffusion coefficient, especially for Ni-based single crystal superalloy which is a multi-component system, whose diffusion coefficient is in the form of a complex matrix. In their creep models, the diffusion coefficient D_L is represented using two quantities, D_o and Q_m, as follows:

$$D_L = D_0 \exp\left(-\frac{Q_m}{RT}\right) \tag{13}$$

$$D_0 = \exp\left(\sum_i x_i ln(D_{o,i}^{Ni})\right) \tag{14}$$

$$Q_m = \sum_i x_i Q_{m,i}^{Ni} \tag{15}$$

where x_i is the mole fraction of solute element i in γ phase, R represents the ideal gas constant and T is the experimental temperature. D_0 and Q_m are the pre-exponential factor and the activation energy, respectively, and both of which can be calculated by summing mole fraction x_i weighted $D_{o,i}^{Ni}$ or $Q_{m,i}^{Ni}$ for all solute elements. $D_{o,i}^{Ni}$ and $Q_{m,i}^{Ni}$ are determined by evaluating diffusion data in individual Ni-i binary alloys, which can be found in the published literature [42–44].

3. Divide-and-conquer self-adaptive learning method

3.1. The main idea and process of the proposed method

The purpose of our research is to establish a creep life prediction model to quantitatively investigate the relationship between creep property and selected material descriptors. However, since the training samples usually contain the alloys with different creep mechanisms, a single model cannot adequately describe all of them. Based on the idea of "divide and conquer", the DCSA automatically separates the alloys according to the different creep mechanisms and is trained on each group independently. As illustrated in Fig. 4, the DCSA is composed of two general stages: automatic division of alloy samples and self-adaptive selection of the optimal model. At the first stage, according to the selected material descriptors, the division algorithm is employed to distinguish the alloy samples with different creep mechanisms and classify them into several homogeneous clusters. Both the similarity of alloys within the same cluster and the difference between the clusters are as large as possible. At the second stage, the selection algorithm is developed to self-adaptively select the optimal regression model for each cluster from the five commonly used regression models. The performance of each model is evaluated using a fitness function. The model with the highest *fitness* value is selected as the optimal model for the corresponding cluster.

The details about two steps of divide-and-conquer self-adaptive learning method are described in the following two sections.

3.2. Automatic division of alloy samples based on K-Means

In this study, K-Means, a clustering algorithm, is used to separate data samples into several clusters, corresponding to different creep mechanisms. As one of the most common unsupervised learning algorithms in data mining, K-Means cluster analysis does not require any prior training data and automatically divides a set into several categories depending on the data distribution [45]. Therefore, in the space of the appropriately selected attributes the alloys with similar creep mechanism can be identified.

K-Means algorithm is a typical distance-based clustering algorithm. Given a set of material samples $S = \{X_1, X_2, \ldots, X_n\}$, where X_i denotes the ith sample. The K-Means algorithm finds the K cluster centers $\{\mu_1, \mu_2, \ldots, \mu_K\}$ which minimize the sum of Cosine distance between each sample X_i and its nearest cluster center μ_j. The Cosine distance between two vectors is presented in Eq. (16).

$$Dis\left(X_i, \mu_j\right) = \frac{\sum_{k=1}^{n} x_{ik} \cdot \mu_{jk}}{\sqrt{\sum_{k=1}^{n} x_{ik}^2} \cdot \sqrt{\sum_{k=1}^{n} \mu_{jk}^2}} \tag{16}$$

where $Dis(X_i, \mu_j)$ denotes the Cosine distance between X_i and μ_j, x_{ik} is the k^{th} feature of the sample i. The closer the distance between the two samples, the higher is the similarity. As a result, the algorithm separates all the samples into K compact and independent clusters.

K-Means firstly chooses K points randomly as the initial centroids $\{\mu_1, \mu_2, \ldots, \mu_K\}$ in the sample space. Then it calculates the distance $Dis(X_i, \mu_j)$ between the sample X_j and each centroid, as shown in Eq. (16), and determines the cluster label λ_j of the sample X_j depending on the nearest cluster center, which is shown in Eq. (17):

$$\lambda_j = \arg\min_{i \in \{1,2,\ldots,K\}} Dis\left(X_i, \mu_j\right) \tag{17}$$

When all samples are assigned with a cluster λ_j, the algorithm recalculates the positions of the K centroids through Eq. (18).

$$\mu_i' = \frac{1}{|\lambda_i|} \sum_{\mathbf{x} \in \lambda_i} X \tag{18}$$

The K-Means repeats calculating the distances and finds new centroids until the centroids no longer change. This produces a separation of the samples into groups from which the metric to be minimized can be calculated. As a result, K-Means clustering partitions the n samples into $K(\leq n)$ sets $C = \{C_1, C_2, \ldots, C_k\}$. The objective

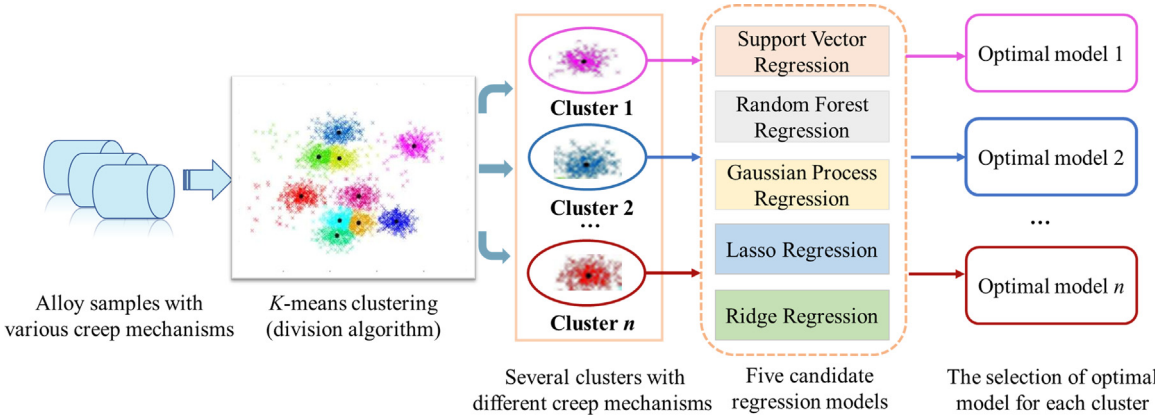

Fig. 4. The procedure of DCSA for modeling the creep rupture life. The input of DCSA is the alloy samples with various creep mechanisms. The division algorithm automatically classifies the samples into several clusters with different creep mechanisms. The selection algorithm is used to self-adaptively choose the optimal regression from five common regression models, i.e. RF, SVR, GPR, LR and RR for each cluster. The final output of DCSA is a group of optimal models for the description of different creep mechanisms for each cluster.

function, as shown in Eq. (17), indicates the similarity of samples within the same cluster. The alloys with similar creep mechanism are thus segregated into clusters. Consequently, the predictive models can then be constructed separately for each homogeneous cluster, rather than the original heterogeneous data set.

3.3. Self-adaptive selection of optimal regression algorithms

Once the subsets of the alloys with similar creep mechanism are identified as described in Section 3.2, the next step of the overall process illustrated in Fig. 4 is to establish the best creep prediction model for each cluster C_k. The problem can be defined as Eq. (19).

$$Y = \begin{cases} f_1(X) & X \in C_1 \\ f_2(X) & X \in C_2 \\ \quad \cdots & \\ f_k(X) & X \in C_k \end{cases} \tag{19}$$

where f_k represents the kth regression model for predicting the alloy samples of the cluster C_k.

From the various available models for capturing the creep rupture life, in this study we select five widely used regression algorithms, including non-linear models such as Random Forest Regression (RF), Support Vector Regression (SVR), and Gaussian Process Regression (GPR) and linear models such as Lasso Regression (LR) and Ridge Regression (RR).

Random Forest Regression (RF) [46] is an ensemble regression model consisting of several decision trees. The bootstrap method is used to extract multiple samples from the original samples and construct decision trees for each bootstrap sample and then use the average of all decision tree predictions as the final prediction result. RF consists of several decision trees and thus the result of RF is easily explainable. In addition, the bootstrap strategy is utilized to overcome the disadvantages of learning on imbalanced problems, i.e. when the number of samples of different classes varies greatly.

Support Vector Regression (SVR) [47] is a nonlinear model extended from the support vector machine (SVM) model. SVR maps input samples from low dimensional space to high dimensional space by kernel function and therefore can effectively tackle nonlinear problems. The structural risk is minimized by introducing the slack variable in the objective function, which enhances its generalization ability, which gives SVR an advantage on datasets with small size and high number of dimensions.

Gaussian Process Regression (GPR) [48] is a Bayesian nonparametric model, which uses lazy learning and a measure of the similarity between points (the kernel function) to predict the value for an unseen point from training data. The prediction provides not just an estimate for that point but also uncertainty information.

Lasso Regression (LR) [49] and Ridge Regression (RR) [49] are simple linear models, which are widely used in prediction of material properties because they can obtain intuitive analytical expressions for the properties of interest. LR and RR can effectively reduce the risk of model overfitting by introducing regularization terms on the basis of simple multiple linear regression.

We define the optimal model via a fitness function to describe the match degree between the creep mechanism and candidate model as shown in Eq. (20).

$$\begin{aligned} &fitness(model, C_i) = fitness(model, \{X_i, Y_i\}) \\ &= 1 - \frac{1}{m}\sum_{j=1}^{m} \frac{\left| model(X_{kj}) - Y_{kj} \right|}{Y_{kj}} \end{aligned} \tag{20}$$

where X_k represents the features of alloy samples in the kth cluster C_i, Y_k represents the creep rupture lives of alloy samples in the kth cluster C_i. Additionally, *model* is one of the candidate algorithms for capturing creep life, and thus $model(X_{ij})$ represents the predicted result. This constructed function quantitatively evaluates how well the model can describe one creep mechanism, that is, the higher *fitness*, the better is the model. On the basis of the fitness function, we further develop the strategy of model-selection for each alloy cluster. As is expressed by Eq. (21), the model with the highest *fitness* is chosen as the optimal model of the cluster C_i.

$$optimal\_model_i = \underset{model \in \{SVR, RF, GPR, LR, RR\}}{argmax} fitness(model, C_i) \tag{21}$$

Through the selection algorithm, the optimal models for each cluster are found, and the ultimate creep model is constructed through embedding these selected regression algorithms, which is shown in Eq. (22).

$$Y = \begin{cases} optimal_model_1(X) & X \in C_1 \\ optimal_model_2(X) & X \in C_2 \\ \quad \cdots & \\ optimal_model_k(X) & X \in C_k \end{cases} \tag{22}$$

4. Experiments

4.1. Experimental dataset

All the data used in this work is collected manually from 14 published patents [50–63]. The creep dataset consists of 266 instances (see Supplementary Table 1), of superalloys of four kinds of test conditions (high-temperature low-stress, high-temperature high-stress, low-temperature low-stress, and low-temperature high-stress) and four generations (1st, 2nd, 3rd, and 4th). The dataset used is available from the resource sharing website [64]. Referring to the analysis in Section 2, 27 relevant material descriptors involving 22 basic physical descriptors and 5 microstructural descriptors, known to correlate with the creep rupture life, were taken as the input. The descriptors are categorized into four different descriptor sets: chemical composition descriptors, test conditions descriptors, heat treatment process descriptors, and microstructural descriptors as shown in Fig. 5. The brief summary of these descriptors is shown in Table 1.

Data pre-processing was necessary before using it for training the model. As shown in Table 1, different material descriptors have large differences in numerical values. For example, the value of shear modulus G ranges from 52.2 GPa to 71.05 GPa while the test temperature distributes from 204 °C to 1800 °C, thus it may produce misleading result during the modeling. To eliminate the influence of these numerical differences on the prediction performance of regression models, the Min-Max normalization method is adopted for preprocessing the descriptors, which is shown in Eq. (23), and thus all the descriptors are mapped to [0, 1].

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{23}$$

where $X_{\min}$ and $X_{\max}$ are the minimum and maximum values for feature X, respectively. In addition, the creep rupture life distributes from 30 h to 6910 h, we treat it by logarithmic scaling as shown in Eq. (24), so that we could conveniently recover the predicted creep rupture life.

$$Y^* = \log(Y) \tag{24}$$

4.2. Experimental setups

The experimental parameters were set as follows. Firstly, we trained the model on the 266 alloy samples collected (see the following section for details). All 266 samples were divided into several clusters by clustering as described in Section 3.2. In each cluster, the alloy samples were divided into the training set and testing set through 10-fold cross-validation to evaluate the generalization

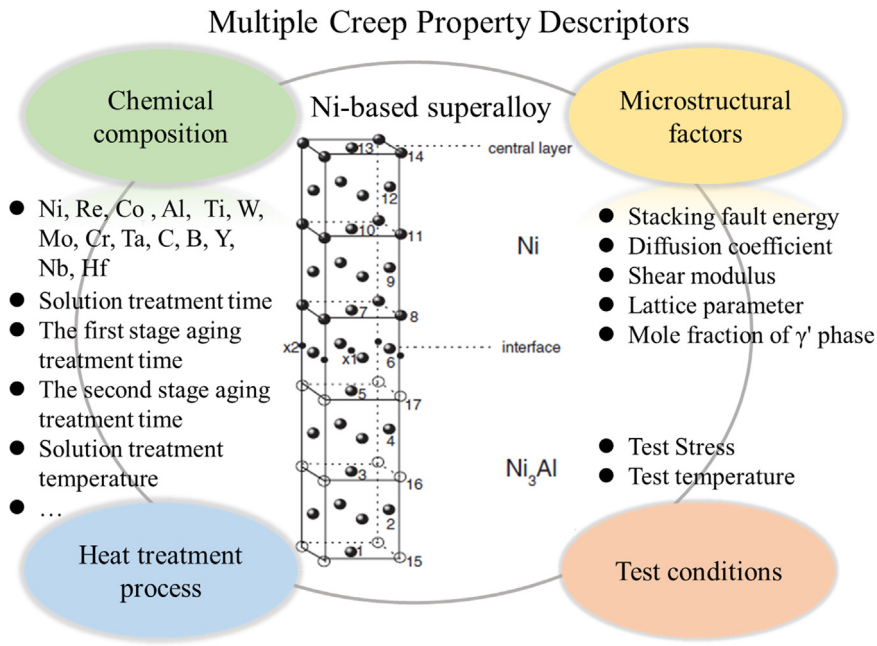

Fig. 5. List of the four different categories of descriptor sets: chemical composition, heat treatment process, microstructural factors, and test conditions.

ability (*fitness* value) of five candidate prediction models for the cluster (SVR, RF, GPR, LR, and RR). In the 10-fold cross-validation on alloy samples of each cluster, 9 folds of samples were used as training dataset *T* for model training and 1 fold as testing dataset *S* for model testing. The overall performance of the candidate model for each cluster is the average of the *fitness* values over all 10 iterations. Furthermore, five single regression prediction models (RF, SVR, GPR, LR, and RR) were used to compare with the DCSA model, and 10-fold cross-validation was also used to evaluate the generalization ability of these models. Finally, eight newly collected alloys were chosen to validate the performance of our model, as described in the following section.

Table 1
Summary of the input material descriptors.

Index	Symbol	Description	Min	Max
1	Ni	Mass percent of Ni	57.41	77.26
2	Re	Mass percent of Re	0	7.5
3	Co	Mass percent of Co	0	12.4
4	Al	Mass percent of Al	4.15	7.5
5	Ti	Mass percent of Ti	0	5
6	W	Mass percent of W	0	18.6
7	Mo	Mass percent of Mo	0	4.21
8	Cr	Mass percent of Cr	0.6	11.73
9	Ta	Mass percent of Ta	0	12
10	C	Mass percent of C	0	0.1
11	B	Mass percent of B	0	0.05
12	Y	Mass percent of Y	0	0.05
13	Nb	Mass percent of Nb	0	2.59
14	Hf	Mass percent of Hf	0	2
15	Stt	Solution treatment time (h)	2	4
16	*1satt*	The first stage aging treatment time (h)	4	5
17	*2satt*	The second stage aging treatment time (h)	16	32
18	*StT*	Solution treatment temperature (°C)	1180	1348
19	*1satT*	The first stage aging treatment temperature (°C)	982	1145
20	*2satT*	The second stage aging treatment temperature (°C)	704	899
21	*T*	Test temperature (°C)	204	1800
22	*S*	Test stress (MPa)	70	759
23	Γ	Stacking fault energy (mJ/m^2)	34.00666	119.4323
24	D_L	diffusion coefficient (m^2/s)	1.41E-25	0.00131
25	*G*	Shear modulus (GPa)	52.20007	71.05828
26	*L*	Lattice parameter (nm)	0.313652	0.362381
27	$X^{\gamma'}$	Mole fraction of γ' phase	0.50187	0.90197

To verify the generalization capability of our DCSA approach, 10-fold cross-validation was performed on 274 (266+8) alloy samples, 9 folds of which were used for model training and 1 fold for model testing. The overall performance of the model is the average of the prediction accuracy over all 10 iterations. In each iteration, we divided the training set into several different clusters through clustering, and selected the model suitable for each cluster from five candidate regression models (i.e. GPR, SVR, RF, LR, and RR). Similarly, in order to prevent overfitting of candidate models, we also conducted 5-fold cross-validation for each cluster to achieve the optimal selection of the model. Regarding the testing set of DCSA model, we calculated the Cosine distance between the test sample and the center of each cluster, so as to find the nearest cluster for each sample, and finally employed the regression model most suitable for the cluster to predict it.

Three widely used evaluation metrics, the root mean square error (*RMSE*) and the mean absolute percent error (*MAPE*) and *R*-square (R^2) were employed to evaluate the predictive performance of regression models. The *RMSE*, *MAPE* and R^2 of DCSA are calculated from the combined prediction results of all clusters and actual measured values.

All the algorithms covered mentioned above are implemented in Python and scikit-learn toolkit [65].

4.3. Experimental results and discussion

In order to confirm the effect of microstructure as a parameter on creep properties, we employed five candidate prediction models (RF, SVR, GPR, LR, and RR) to conduct predictive analysis on the creep dataset without and with these descriptors, respectively. The average *MAPE* and *RMSE* of 10 iterations of five candidate models on the dataset without and with these microstructural descriptors are shown in Table 2. We can observe from Table 2 that the prediction accuracy of the five candidate models on the creep dataset with five microstructural descriptors is a slightly better than that without these microstructural descriptors. Actually, although a minor improvement is obtained, we

Table 2
The comparison of prediction performance (with standard deviation) of five models on the creep dataset without and with five microstructural descriptors.

Candidate models	With microstructural descriptors		Without microstructural descriptors	
	RMSE	*MAPE*	*RMSE*	*MAPE*
RF	0.5229±0.5068	0.0599±0.0614	0.5383±0.5074	0.0618±0.0565
SVR	0.7303±1.0782	0.0524±0.0487	0.7422±1.0665	0.0539±0.0475
GPR	0.6619±0.8223	0.0759±0.0794	0.6798±0.8294	0.0788±0.0806
LR	1.2504±1.6798	0.0896±0.0961	1.2591±1.6785	0.0899±0.0962
RR	0.9247±1.1514	0.0693±0.0609	0.9170±1.1510	0.0697±0.0860

can effectively predict creep properties using only the information on the composition, structure, process, and test conditions.

All the 27 features (descriptors) are considered to be associated with creep rupture life. Therefore, to further identify the importance of the features, we performed Random Forest Regression (RF) on the creep dataset with the five calculated microstructural descriptors. The resultant average feature importance for each feature is shown in Fig. 6. It is found that heat treatment process parameters including the 2nd stage aging treatment temperature 2*satT* and the 2nd stage aging treatment time 2*satt* play the most important role in creep rupture life, followed by test conditions including the test stress *S* and the test temperature *T*. This result may be due to the fact that the numerical values of the test conditions of our collected alloy samples did not change as much as the heat treatment process parameters. Furthermore, as for the alloy composition, the mass percent of alloy elements Cr, Ta, Re and Mo are important factors to the creep property. Encouragingly, we also found that the microstructural descriptors including stacking fault energy τ, shear modulus *G*, lattice parameter *L*, and mole fraction of γ' phase also have impact on the creep rupture life of the alloy.

Based on the 27 features, the K-Means clustering was used to classify the alloy samples with different creep mechanisms. *K* value of the division algorithm is set at 8 after trial and error. As a result, all the alloy samples were divided into eight homogenous clusters, and the number of samples in each cluster is shown as Fig. 8. The number of alloy samples in the cluster 3 is the largest, 63, and the number of alloy samples in the clusters 1, 2, 4, 5, and 7 are 38, 36, 36, 44 and 21, respectively. The number of alloy samples in the clusters 6 and 8 (10 and 18, respectively) is relatively small, which may cause the overfitting of the predictive model on these two clusters. Therefore, the predictive accuracy of the model on each cluster is obtained by means of 10-fold cross-validation. Principal component analysis (PCA) was used to extract the dominating components from the original input features. We chose two major components (PC1 and PC2, with relative weightings of 0.3830 and 0.2119, respectively) to substitute the initial dimensional representation, in order to intuitively visualize the distribution of sample points in each cluster by means of two-dimensional scatter diagram, as shown in Fig. 7. Notably, we can observe that the clusters 2, 6, and 8 are overlapping, which is a result of the projection onto the PC1-PC2 plane. Furthermore, Fig. 7 also shows that except the clusters 1 and 8, the remaining 6 clusters are compact. For the two abnormal clusters 1 and 8, we analyzed the composition, heat treatment process and test conditions in more details and found that the content W and Ta, test temperature and stress of several samples fluctuated in a wide range, which may affect performance of the predictive model on these two clusters and produce large deviations in the predicted results.

Next, we analyzed whether the division algorithm can distinguish different creep mechanisms. As discussed in Section 3, superalloys of different generations and test conditions have significant difference in creep mechanisms, and thus we can verify whether our division algorithm accurately identifies different creep mechanisms based on the descriptors in each cluster. Generally, in Ni-based single crystal superalloys, the content of Re element is related to the generation of the alloy. The first generation of superalloys does not contain Re, the second generation contains about 3 wt% Re, and the third and fourth generations contain about 6 wt% Re or more. Hence, we can observe from Fig. 9(a) and Table 3 that the content of Re in the cluster 1 is close to zero, indicating that the alloys in this cluster all belong to the first-generation. The content of Re in clusters 2, 7, and 8 ranges from 0 wt% to 3 wt%, which are 1.7447 wt%, 2.9333 wt%, and 1.6667 wt% respectively, indicating that the alloys in these clusters belong to the

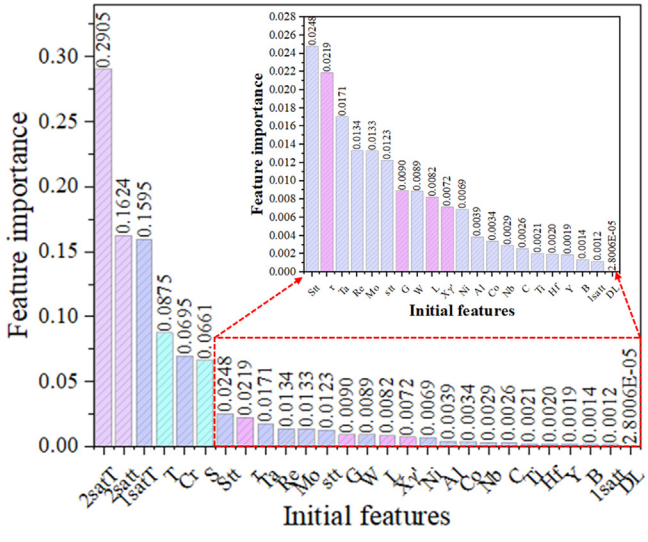

Fig. 6. The ranking of feature importance of initial input variables.

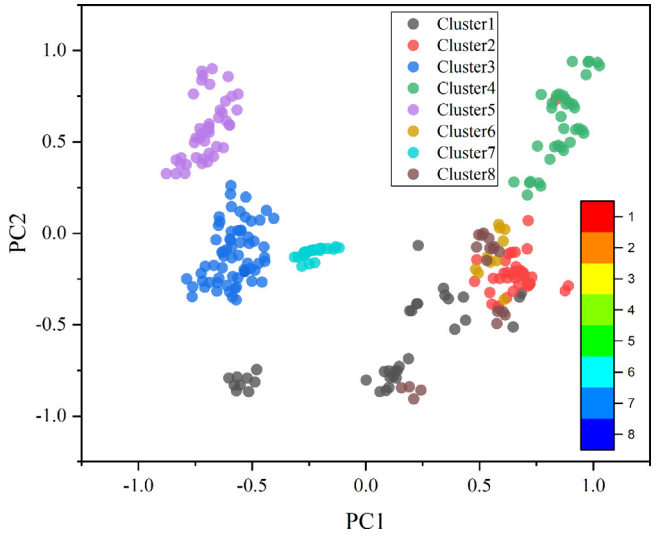

Fig. 7. Two-dimensional scattering distribution of clustering results. The *X*-coordinate and *Y*-coordinate represent principal components PC1 and PC2, respectively.

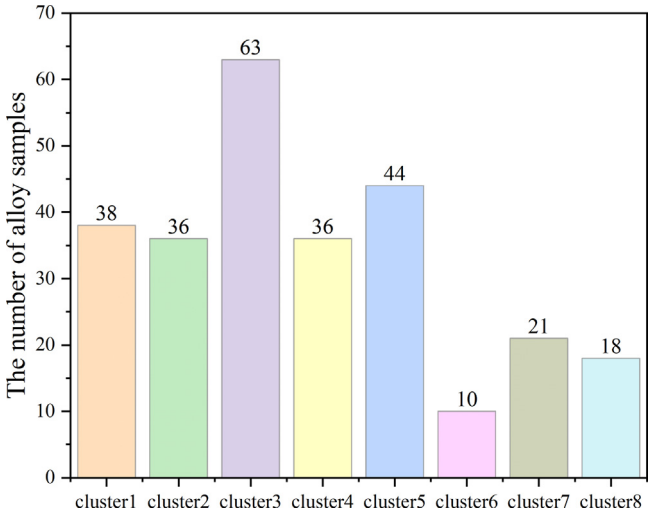

Fig. 8. Number of samples in each cluster.

second-generation superalloys. Similarly, the alloys in the clusters 3, 5 and 6 are the third-generation alloys with the content of Re (3.4032 wt%, 3.4091 wt% and 3.02 wt%) ranging from 3 wt% to 6 wt%. The alloys in the cluster 4 are fourth-generation superalloys, having the content of Re of about 6 wt%. Fig. 9(b), (c) and Table 3

Table 3
The distinguishable features (the content of Re, test stress and test temperature) of eight clusters.

Cluster	The mean content of Re (wt%)	The mean test stress (MPa)	The mean test temperature (°C)
cluster1	0	371.9826	850.6842
cluster2	1.7447	137.2061	1040
cluster3	3.4032	331.3333	977.5556
cluster4	6.3000	193.3333	1050
cluster5	3.4091	256.6364	1027.2728
cluster6	3.0200	248.4000	1800
cluster7	2.9333	202.1429	950
cluster8	1.6667	233.7833	1010.5000

demonstrate that the mean test conditions of the alloys in the clusters 2 and 4 belong to high temperature and low stress (1040 °C/137.2061 MPa, 1050 °C/193.3333 MPa), while the clusters 1, 3, and 7 (850.6842 °C/371.9826 MPa, 977.5556 °C/331.3333 MPa and 950 °C/202.1429 MPa) contains medium temperature and high stress alloys. The mean test conditions (1027.2728 °C/256.6364 MPa, 1010.5 °C/233.7833 MPa) of the alloys in the clusters 5 and 8 belong to high temperature and medium stress. In addition, the division algorithm differentiates the alloys prepared under special conditions in the dataset, and the results show that the cluster 6 belongs to ultrahigh temperature and medium stress, which has the large difference from the other clusters. To sum up, the division algorithm can effectively

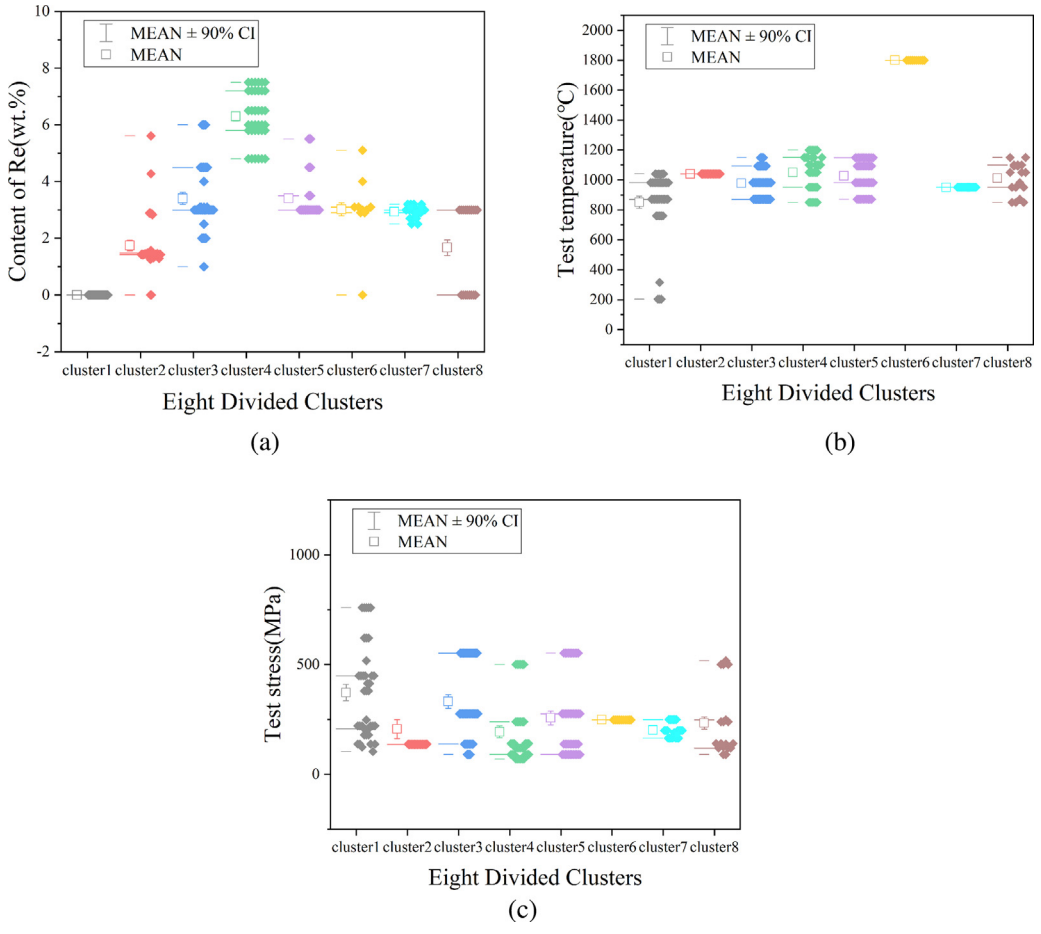

Fig. 9. The box plot of the content of Re and test conditions (test temperature and test stress) in the eight divided clusters. (a)–(c) reveal the mean values with 90% confidence coefficient of the content of Re, test temperature and test stress of the eight clusters, respectively. Different colored dots represent alloy samples in each cluster.

distinguish the alloys with different creep mechanisms which improves the accuracy of creep prediction model.

Furthermore, the optimal model-selection algorithm was used to find an optimal model among Support Vector Regression (SVR), Random Forest Regression (RF), Gaussian Process Regression (GPR), Ridge Regression (RR), and Lasso Regression (LR) for each cluster. As described in Section 3.3, the *fitness* values of the five candidate regression models for each cluster are the results of 10-fold cross-validation. The prediction performance of these models in each cluster is shown in Tables 4 and 5. Next, taking the cluster 1 as an example, the *fitness* values (with standard deviation) of five candidate regression models for the cluster are 0.9242, 0.8812, 0.8618, 0.8648 and 0.8554, respectively. Obviously, for this cluster, the average *fitness* value of SVR is higher than those of RR, GPR, RF and LR. When the *fitness* values of several models for a cluster are close, the optimal model can be further selected for the cluster based on the smaller standard deviation. According to this criterion, we can select the prediction model with high *fitness* value and good stability for each cluster. As shown in Table 6, the SVR also has best results and is selected as the optimal model for the clusters 6 and 8. The RR is selected for the cluster 2, the RF is selected for the clusters 4, 5, and 7, and the GPR is chosen for the cluster 3. Furthermore, the maximum and minimum *fitness* values of the five candidate predicted models on each cluster are shown in Fig. 10. It can be seen that there are some differences in the adaptability of different models on each cluster, especially for cluster 1,

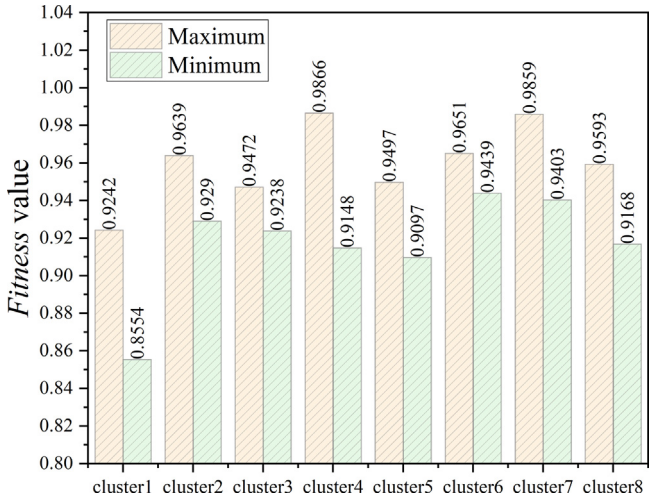

Fig. 10. The maximum and minimum *fitness* values of each model on eight clusters with different creep rupture mechanisms.

cluster 4, cluster 5, cluster 7 and cluster 8. All these prove that the self-adaptive selection of the model is relatively effective.

In order to further demonstrate the validity of our divide-and-conquer self-adaptive learning method, six regression models including RF, SVR, GPR, RR, LR and our model DCSA were all employed to predict the creep rupture life. Table 7 lists the average and standard deviation values (10-fold cross-validation) for *RMSE* and *MAPE*, and R^2 of six different prediction models. The lower the average value of *RMSE* and *MAPE* is, the better the prediction effect of the model is. Therefore, we can observe that our DCSA model gives the lowest *RMSE* (0.3839), *MAPE* (0.0003) values and the highest R^2 value (0.9176) and is thus superior to the other five single regression models in prediction accuracy. The performance of model training is shown in Fig. 11 with measured creep rupture life (true value) on the horizontal axis and predicted creep rupture life (predicted value) on the vertical axis. The predicted creep rupture life values generated by the above six models are plotted as a function of the measured. The more closely the plots align along the 45_diagonal line is, the better the prediction is. We can see that the DCSA model produces the best performance, as shown in Fig. 11(f). Furthermore, it can be observed that the number of serious outliers for the DCSA model is far less than that of any of the other five models. In Fig. 11(f), the different colors represent the prediction for different clusters. We can observe that the optimal model selected for each cluster achieves good fit, especially for the clusters 4, 6, 7 and 8. However, the model did not yield good fit for several points of the cluster 1. Therefore, we carefully examined the input features of these clusters, including the chemical composition and test temperature and test stress and

Table 4
Prediction accuracy (*fitness* value with standard deviation) of five candidate regression models for clusters 1, 2, 3 and 4.

Model	*Fitness* value (10-fold cross-validation)			
	cluster1	cluster2	cluster3	cluster4
SVR	**0.9242±0.0416**	0.9473±0.0248	0.9323±0.0470	0.9720±0.0212
RF	0.8812±0.0923	0.9432±0.0393	0.9321±0.0435	**0.9866±0.0278**
GPR	0.8618±0.1158	0.9496±0.0467	**0.9472±0.0434**	0.9148±0.0581
RR	0.8648±0.1023	**0.9639±0.0189**	0.9238±0.0397	0.9160±0.0433
LR	0.8554±0.1337	0.9290±0.0368	0.9314±0.0478	0.9296±0.0412

Table 5
Prediction accuracy (*fitness* value with standard deviation) of five candidate regression models for clusters 5, 6, 7 and 8.

Model	*Fitness* value (10-fold cross-validation)			
	cluster5	cluster6	cluster7	cluster8
SVR	0.9468±0.0362	**0.9651±0.0277**	0.9806±0.0128	**0.9593±0.0303**
RF	**0.9497±0.0338**	0.9612±0.0402	**0.9859±0.0105**	0.9349±0.0754
GPR	0.9404±0.0455	0.9615±0.0384	0.9761±0.0126	0.9168±0.0859
RR	0.9462±0.0369	0.9439±0.0309	0.9500±0.0367	0.9511±0.0367
LR	0.9097±0.0562	0.9611±0.0518	0.9403±0.0243	0.9455±0.0499

Table 6
Selected optimal regression models for eight clusters.

Clusters	Selected optimal model	*Fitness*
1	SVR	0.9242
2	RR	0.9639
3	GPR	0.9472
4	RF	0.9866
5	RF	0.9497
6	SVR	0.9651
7	RF	0.9859
8	SVR	0.9593

Table 7
Average and standard deviation values (10-fold cross-validation) for *RMSE* and *MAPE* of six different prediction models.

Prediction models	*RMSE*		*MAPE*		R^2
	Ave	Dev	Ave	Dev	
RF	0.5229	0.5068	0.0599	0.0614	0.7102
SVR	0.7303	1.0782	0.0524	0.0487	0.5976
GPR	0.6619	0.8223	0.0759	0.0794	0.6338
LR	1.2504	1.6798	0.0896	0.0961	0.3742
RR	0.9247	1.1514	0.0693	0.0609	0.4895
DCSA (Our model)	**0.3839**	–	**0.0003**	–	**0.9176**

*Ave = Average, Dev = Deviation.

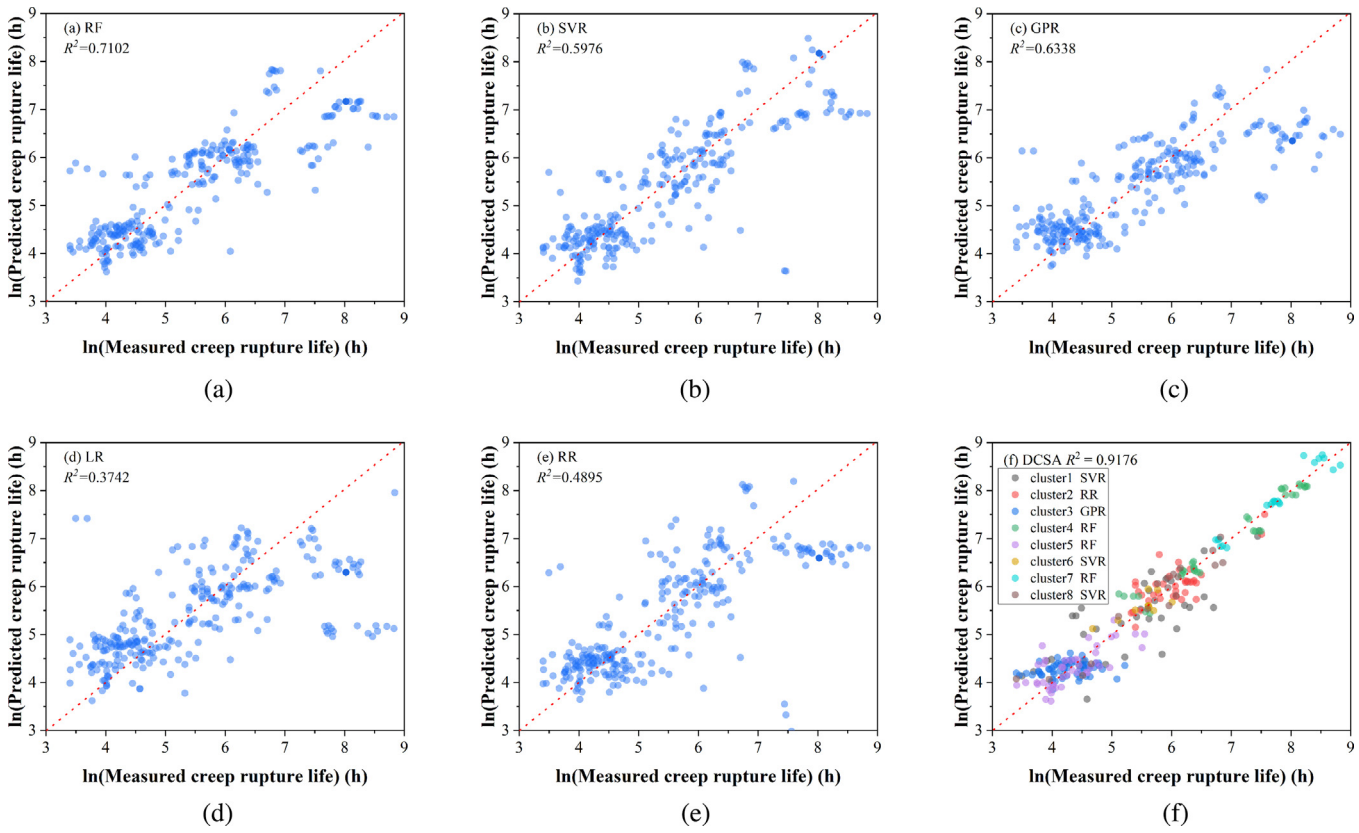

Fig. 11. Predictive performance of the six machine learning models on the creep dataset. (a) Random Forest (RF); (b) Support vector regression (SVR); (c) Gaussian process regression (GPR); (d) Lasso Regression (LR); (e) Ridge Regression (RR); (f) our method (DCSA).

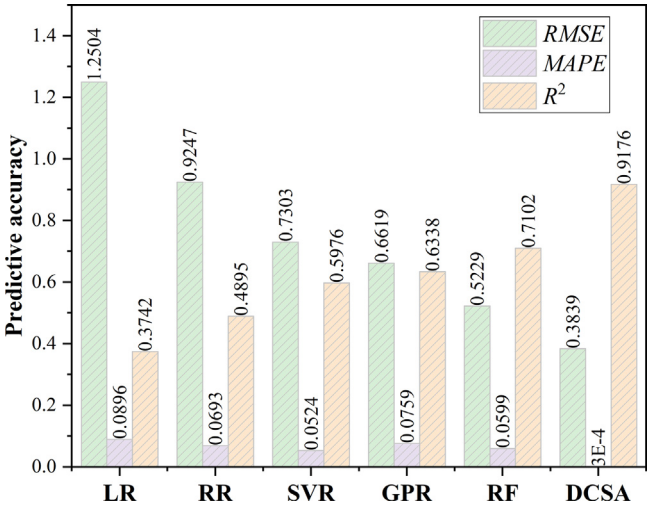

Fig. 12. The *RMSE, MAPE* and R^2 of different machine learning models on the creep dataset.

discovered that the content of the elements W, Ta and Co, test temperature and test stress values of these sample points fluctuate within a larger range compared to those of the samples in the other clusters. Furthermore, as can be seen from Fig. 7, since the sample points of the cluster 1 are widely scattered, all five candidate models produced lower than average *fitness* values on this cluster. The values of *RMSE, MAPE* and R^2 of six regression models are depicted in Fig. 12. We can see that compared to the other five baseline models, the DCSA model has the lowest *RMSE, MAPE* values of 0.3839 and 0.0003, respectively, and the highest R^2 of 0.9176, indicating the superior performance of the model.

To validate the accuracy of our DCSA model, we applied it to the newly collected dataset of creep properties, previously unseen by the model. The validation dataset is from the four published patents [66–68,53], which are not included in our training set with 266 instances. The validation dataset consists of 8 instances (#1 (1040 °C, 137.2 MPa), # 2 (1040 °C, 137.2 MPa), #3 (204 °C, 759 MPa), #4 (315 °C, 448.5 MPa), #5 (900 °C, 392 MPa), #6 (1100 °C, 137 MPa), #7(982 °C, 248.4 MPa) and #8(982 °C, 248.4 MPa)) (see Supplementary Table 2). There are two main reasons why we choose these instances for experimental validation. One is that the range of chemical composition of these eight instances varies significantly and we can test the sensitivity of the DCSA model to different elemental components, especially Co (0.99–12.4), Re (0–5.61), and Cr (4.27–10.6), whose contents have a great influence on creep and fatigue properties. The other is that these eight instances involve low-temperature high-stress and low-temperature medium-stress test conditions that are different from the training set with 266 instances.

We firstly employed Thermo-Calc method and basic material formulas discussed in Section 2 to calculate the five microstructural descriptors that are linked to creep properties for the newly designed alloys, as shown in Table 8.

Based on the 27 input features, we calculated the Cosine distances between eight new samples and the cluster centers of 8 clusters constructed during training to find the optimal cluster to which these samples belong, and finally selected the most suitable prediction model, as shown in Table 9. It can be observed that #1, #2 and #5 select the optimal predictive model SVR of

Table 8
Calculated values of five microstructural descriptors including stacking fault energy Γ, diffusion coefficient D_L, shear modulus G, lattice parameter L and mole fraction of γ' phase $X^{\gamma'}$.

Instance	Γ (mJ/m^2)	D_L (m^2/s)	G (GPa)	L (nm)	$X^{\gamma'}$	Creep life (h)
#1 [67]	61.0643	4.5170E−21	53.9646	0.3508	0.7521	454.3
#2 [67]	59.7776	2.2470E−21	57.5799	0.3576	0.5464	211.1
#3 [66]	41.0473	1.8070E−21	69.0491	0.3493	0.6024	256.2
#4 [66]	41.3422	1.8000E−21	69.2008	0.3497	0.5966	277.1
#5 [68]	40.9272	5.9446E−21	66.8093	0.3494	0.7190	652
#6 [68]	39.7587	6.1845E−18	62.6643	0.3507	0.7090	351
#7 [53]	45.7544	3.0460E−15	60.3920	0.3554	0.6604	199.2
#8 [53]	39.2383	1.0460E−15	61.1557	0.3602	0.5654	129.2

Table 9
Alloy clusters and corresponding predictive models selected for eight newly collected alloy samples.

Instance	Alloy clusters	Selected predictive model
#1	cluster2	RR
#2	cluster2	RR
#3	cluster1	SVR
#4	cluster1	SVR
#5	cluster2	RR
#6	cluster8	SVR
#7	cluster1	SVR
#8	cluster1	SVR

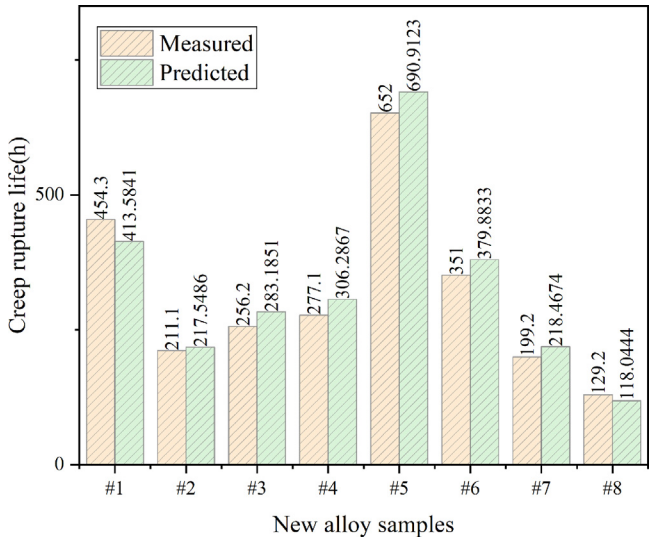

Fig. 13. Comparison of measured and predicted values of creep rupture life for eight newly collected alloys.

the cluster 2, #3, #4, #7 and #8 select the optimal predictive model SVR of the cluster 1, and #6 selects the optimal predictive model SVR of the cluster 8. As shown in Fig. 13, the difference between the measured and predicted values of the model on these eight alloy examples is found within acceptable range, that validates our model DCSA.

Finally, in order to evaluate the overall performance of our model DCSA, we applied it to all 274 (266+8) alloy samples (see Section 4.2 for experimental setups). As shown in Fig. 14, the average *RMSE*, *MAPE* (with standard deviation) and R^2 of the model are 0.3026 (0.0499), 0.0499 (0.0513) and 0.8268, respectively, indicating both the good predictive performance of our DCSA model and its stability with respect to the unseen data.

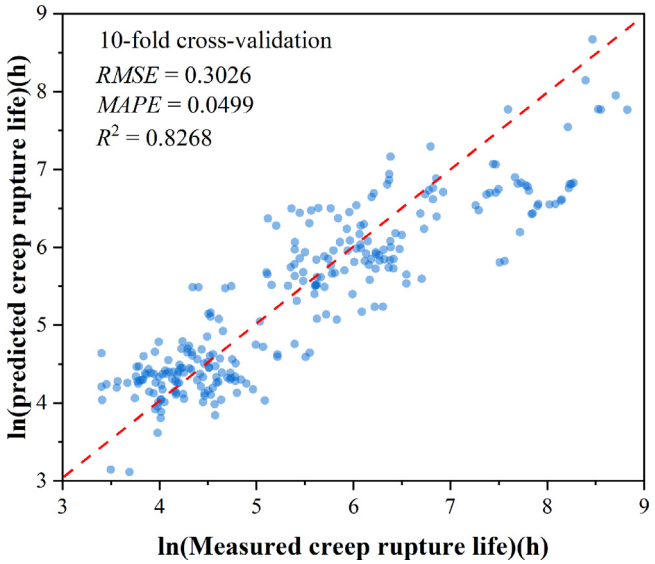

Fig. 14. Overall predictive performance of our model DCSA on all 274 alloy samples (10-fold cross-validation).

5. Conclusions

The divide-and-conquer self-adaptive (DCSA) machine learning method has been developed to accelerate the prediction of creep rupture life of Ni-based single crystal superalloys. Five microstructural parameters related to creep properties (stacking fault energy, lattice parameter, mole fraction of γ' phase, diffusion coefficient and shear modulus) were calculated and introduced to describe the creep behavior. The DCSA method consists of two steps. First, it utilizes the division algorithm to automatically cluster the alloys with different creep mechanisms and then self-adaptively identifies the optimal regression model for each cluster. Through extensive testing we confirmed that the division algorithm can effectively distinguish different classes of superalloys and select the optimal model for further prediction. Comparative study of the DCSA and five state-of-the-art models shows that the R^2 values from RF, SVR, GPR, LR and RR are 0.7102, 0.5976, 0.6338, 0.3742 and 0.4895 respectively, whereas that of the DCSA is 0.9176, demonstrating the effectiveness of the strategy of divide and conquer in predicting creep rupture life. For the newly acquired eight alloy samples, the error between the predicted values and the experimental measured values is within an acceptable range (6.4486 h−40.7159 h), further confirming the effectiveness of our DCSA method.

Overall, we developed an effective machine learning approach for creep rupture life prediction of Ni-based single crystal superalloys. Our method accelerates the prediction of creep rupture life dramatically, compared to exhaustive experimental testing, based only on the initial information of chemical composition, test conditions, heat treatment and microstructural parameters, which allows to reduce time and cost during alloy design. We also note that this approach can be easily extended to take into account other effects relevant for creep, e.g. activation energy, lattice misfit, etc. and this will be the focus of the future work.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgment

This work is supported by the National Key R&D Program of China (Nos. 2017YFB0701502, 2017YFB0701501, 2017YFB0701503), Open Project of the State Key Laboratory of Advanced Special Steel,

Shanghai University, China (No. SKLASS2018-01) and the Project of the State Key Laboratory of Advanced Special Steel, Shanghai University, China (No. SKLASS2019-Z023) and the Science and Technology Commission of Shanghai Municipality (No. 19DZ2270200). We appreciate the High Performance Computing Center of Shanghai University, and Shanghai Engineering Research Center of Intelligent Computing System (No. 19DZ2252600) for providing the computing resources and technical support.

Author contributions

Yue Liu, Maxim Avdeev, Xiaogang Lu, Chongyu Wang, Tao Yu and Siqi Shi conceived and designed the theoretical calculations and experiments. Yue Liu, Junming Wu, Zhichao Wang and Siqi Shi carried out all the theoretical calculations and analyzed the data. Yue Liu, Junming Wu and Zhichao Wang performed and analyzed all the experiments. All authors discussed the results and prepared and revised the manuscript.

Data and code availability

The authors declare that the main data supporting the finding of this study are available within the article and its Supplementary Information files. Extra data are available from the corresponding author upon reasonable request. All source codes for the experiments are implemented in Python and uploaded to the author's github repository: https://github.com/wujunming1/mla-shu.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.actamat.2020.05.001.

References

[1] H. Long, S. Mao, Y. Liu, et al., Microstructural and compositional design of Ni-based single crystalline superalloys—a review, J. Alloys Compd. 743 (2018) 203–220.
[2] T.M. Pollock, S. Tin, Nickel-based superalloys for advanced turbine engines: chemistry, microstructure and properties, J. Propul. Power 22 (2) (2006) 361–374.
[3] L. Cui, J. Yu, J. Liu, et al., The creep deformation mechanisms of a newly designed nickel-base superalloy, Mater. Sci. Eng. A 710 (2018) 309–317.
[4] K Rajan, Materials informatics, Mater. Today 8 (10) (2005) 38–45.
[5] Y.Y. Dang, X.B. Zhao, Y. Yuan, et al., Predicting long-term creep-rupture property of Inconel 740 and 740H, Mater. High Temp. 33 (1) (2016) 1–5.
[6] J. Bolton, Reliable analysis and extrapolation of creep rupture data, Int. J. Press. Vessels Pip. 157 (2017) 1–19.
[7] D.W. MacLachlan, D.M. Knowles, Modelling and prediction of the stress rupture behaviour of single crystal superalloys, Mater. Sci. Eng.: A 302 (2) (2001) 275–285.
[8] L. Feng, K. Zhang, G. Zhang, et al., Anisotropic damage model under continuum slip crystal plasticity theory for single crystals, Int. J. Solids Struct. 39 (20) (2002) 5279–5293.
[9] S.C. Prasad, .K..R Rajagopal, I.J. Rao, A continuum model for the anisotropic creep of single crystal nickel-based superalloys, Acta Mater. 54 (6) (2006) 1487–1500.
[10] I.N. Vladimirov, S. Reese, G Eggeler, Constitutive modelling of the anisotropic creep behaviour of nickel-base single crystal superalloys, Int. J. Mech. Sci. 51 (4) (2009) 305–313.
[11] B. Fedelich, A. Epishin, T. Link, et al., Experimental characterization and mechanical modeling of creep induced rafting in superalloys, Comput. Mater. Sci. 64 (2012) 2–6.
[12] Y.K. Kim, D. Kim, H.K. Kim, et al., An intermediate temperature creep model for Ni-based superalloys, Int. J. Plast. 79 (2016) 153–175.
[13] D.C. Elton, Z. Boukouvalas, M.S. Butrico, et al., Applying machine learning techniques to predict the properties of energetic materials, Sci. Rep. 8 (1) (2018) 9059.
[14] Y. Liu, T.L. Zhao, G. Yang, et al., The onset temperature (T_g) of As_xSe_{1-x} glasses transition prediction: a comparison of topological and regression analysis methods, Comput. Mater. Sci. 140 (2017) 315–321.
[15] Y. Liu, J.M. Wu, G. Yang, T.L. Zhao, S.Q. Shi, Predicting the onset temperature (T_g) of Ge_xSe_{1-x} glass transition: a feature selection based two-stage support vector regression method, Sci. Bull. 64 (16) (2019) 1195–1203.
[16] J.M. J Granda, L. Donina, V. Dragone, et al., Controlling an organic synthesis robot with machine learning to search for new reactivity, Nature 559 (7714) (2018) 377.
[17] J.C. Snyder, M. Rupp, K. Hansen, et al., Finding density functionals with machine learning, Phys. Rev. Lett. 108 (25) (2012) 253002.
[18] Y. Liu, T. Zhao, W. Ju, et al., Materials discovery and design using machine learning, J. Materiom. 3 (3) (2017) 159–177.
[19] V. Venkatesh, Rack, A neural network approach to elevated temperature creep–fatigue life prediction, Int. J. Fatigue 21 (3) (1999) 225–234.
[20] Y.S. Yoo, C.Y. Jo, C.N. Jones, Compositional prediction of creep rupture life of single crystal Ni base superalloy by Bayesian neural network, Mater. Sci. Eng. A 336 (1–2) (2002) 22–29.
[21] A. Royer, P. Bastie, M. Veron, In situ determination of γ' phase volume fraction and of relations between lattice parameters and precipitate morphology in Ni-based single crystal superalloy, Acta Mater. 46 (15) (1998) 5357–5368.
[22] T. Ning, T. Sugui, Y. Huajin, et al., Deformation mechanisms and analysis of a single crystal nickel-based superalloy during tensile at room temperature, Mater. Sci. Eng.: A 744 (2019) 154–162.
[23] M.E. Kassner, M.T. Perez-Prado, Five-power-law creep in single phase metals and alloys, Prog. Mater. Sci. 45 (1) (2000) 1–102.
[24] X. Wu, C. Wang, Density functional theory study of the thermodynamic and elastic properties of Ni-based superalloys, J. Phys. Condens. Matter 27 (29) (2015) 295401.
[25] J.X. Zhang, C. Wang J, H. Harada, et al., The effect of lattice misfit on the dislocation motion in superalloys during high-temperature low-stress creep, Acta Mater. 53 (17) (2005) 4623–4633.
[26] T.M. Pollock, A.S. Argon, Creep resistance of CMSX-3 nickel base superalloy single crystals, Acta Metall. Mater. 40 (1) (1992) 1–30.
[27] T.M. Pollock, A.S. Argon, Directional coarsening in nickel-base single crystals with high volume fractions of coherent precipitates, Acta Metall. Mater. 42 (6) (1994) 1859–1874.
[28] F. Wu, H. Murakami, A Suzuki, Development of an iridium–tantalum modified aluminide coating as a diffusion barrier on nickel-base single crystal superalloy TMS-75, Surf. Coat. Technol. 168 (1) (2003) 62–69.
[29] J.X. Zhang, T. Murakumo, Y. Koizumi, et al., Interfacial dislocation networks strengthening a fourth-generation single-crystal TMS-138 superalloy, Metall. Mater. Trans. A 33 (12) (2002) 3741–3746.
[30] P. Shi, A. Engström, L. Höglund, et al., Thermo-Calc and DICTRA enhance materials design and processing, Materials Science Forum 475 (2005) 3339–3346 Trans Tech Publications.
[31] A. Müller, I. Roslyakova, M. Sprenger, et al., MultOpt++: a fast regression-based model for the development of compositions with high robustness against scatter of element concentrations, Model. Simul. Mater. Sci. Eng. 27 (2) (2019) 024001.
[32] R. Rettig, A. Heckl, S. Neumeier, et al., Verification of a commercial CALPHAD database for Re and Ru containing nickel-base superalloys, Defect Diffus. Forum 289 (2009) 101–108 Trans Tech Publications.
[33] C. Ritter N, E. Schesler, A. Müller, et al., On the influence of Ta and Ti on heat-–treatability and γ/γ'–partitioning of high W containing Re–free nickel–based superalloys, Adv. Eng. Mater. 19 (8) (2017) 1700150.
[34] E.H. Copland, N.S. Jacobson, F.J. Ritzert, Computational Thermodynamic Study to Predict Complex Phase Equilibria in the Nickel-Base Superalloy René N6. Report No. NASA/TM-2001- 210897 (2001) 1–42.
[35] S. Sui, H. Tan, J. Chen, et al., The influence of Laves phases on the room temperature tensile properties of Inconel 718 fabricated by powder feeding laser additive manufacturing, Acta Mater. 164 (2019) 413–427.
[36] K.J. Niklas, Voigt and Reuss models for predicting changes in Young's modulus of dehydrating plant organs, Ann. Bot. 70 (4) (1992) 347–355.
[37] Z. Guo, A.P. Miodownik, N. Saunders, et al., Influence of stacking-fault energy on high temperature creep of alpha titanium alloys, Scripta Mater. 54 (12) (2006) 2175–2178.
[38] S.L. Shang, D.E. Kim, C.L. Zacherl, et al., Effects of alloying elements and temperature on the elastic properties of dilute Ni-base superalloys from first-principles calculations, J. Appl. Phys. 112 (5) (2012) 053515.
[39] S.L. Shang, C.L. Zacherl, H.Z. Fang, et al., Effects of alloying element and temperature on the stacking fault energies of dilute Ni-base superalloys, J. Phys.: Condens. Matter 24 (50) (2012) 505403.
[40] N. Saunders, Z. Guo, X. Li, et al., Modelling the material properties and behaviour of Ni-based superalloys, Superalloys 2004 (2004) 849–858.
[41] Z. Zhu, H. Basoalto, N. Warnken, et al., A model for the creep deformation behaviour of nickel-based single crystal superalloys, Acta Mater. 60 (12) (2012) 4888–4900.
[42] C.E. Campbell, W.J. Boettinger, U.R. Kattner, Development of a diffusion mobility database for Ni-base superalloys, Acta Mater. 50 (4) (2002) 775–792.
[43] X.J. Liu, H.H. Hu, J.J. Han, et al., Assessment of the diffusional mobilities in fcc Ni–Nb and fcc Ni–Mo alloys, Calphad 38 (2012) 140–145.
[44] X.J. Liu, Y. Lin J, Y. Lu, et al., Assessment of the atomic mobility for the fcc phase of Ni–Co–X (X= Re and Ru) system, Calphad 45 (2014) 138–144.
[45] G. Tzortzis, A. Likas, The MinMax k-Means clustering algorithm, Pattern Recognit. 47 (7) (2014) 2505–2516.
[46] A. Liaw, M. Wiener, Classification and regression by randomForest, R News 2 (3) (2002) 18–22.
[47] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004) 199–222.
[48] J. Quiñonero-Candela, C.E. Rasmussen, A unifying view of sparse approximate Gaussian process regression, J. Mach. Learn. Res. 6 (Dec) (2005) 1939–1959.
[49] Art B. Owen, A robust hybrid of lasso and ridge regression, Contemp. Math. 443 (7) (2007) 59–72.
[50] M. Konter, M. Newnham, C. Tonnes, Nickel-Base Superalloy, US Patent, 5,888,451 (1999) 1-10.

[51] E.W. Ross, C.S. Wukusick, W.T. King, Nickel-Based Superalloys for Producing Single Crystal Articles Having Improved Tolerance to Low Angle Grain Boundaries, US Patent, 5, 399, 313 (1995) 1-16 .
[52] C.S. Wukusick, L. Buchakjian Jr, R. Darolia, Heat Treatment for Nickel-Base Superalloys, US Patent, 5,100,484 (1992) 1-12 .
[53] F.A. Schweizer, Single Crystal Nickel-Base Super Alloy, US Patent, 4,765,850 (1988) 1–6.
[54] F.A. Schweizer, X. Nguyen-Dinh, Single Crystal Nickel-Base Superalloy for Turbine Components, US Patent, 5,077,004 (1991)1-6.
[55] S.W. Shaw, Nickel-Base Superalloys, US Patent , 4,207,098 (1980) 1-6.
[56] N. Das, Nickel-Based Superalloy, US Patent, 5,925,198 (1999) 1-10.
[57] M. Yamazaki, T. Yamagata, H. Harada, Nickel-Base Single Crystal Superalloy and Process for Production Thereof, US Patent, 4,707,192, (1987) 1-12.
[58] G.E. Maurer, W.J. Boesch, J.M. Theret, Nickel Base Alloy, US Patent, 4,629,521 (1986) 1-10.
[59] D.N. Duhl, Olson W.E.Heat Treated Superalloy Single Crystal Article and Process, US Patent, 4,209,348 (1980) 1–10.
[60] F.A. Schweizer, D.N. Duhl, Single Crystal Nickel Superalloy, U.S. Patent 4,222,794, (1980)1-8.
[61] D.N. Duhl, X. Nguyen-Dinh, Single Crystal Nickel Superalloy, US Patent, 4,371,404 (1983)1-8.
[62] D.N. Duhl, A.D. Cetel, Advanced High Strength Single Crystal Superalloy Composition, US Patent, 4,719,080 (1988) 1–8.
[63] S. Chin, D,N. Duhl, High Strength Single Crystal Superalloys, US Patent, 4,908,183 (1990)1-6.
[64] https://github.com/wujunming1/mla-shu(accessed: September 2019).
[65] F. Pedregosa, G. Varoquaux, A. Gramfort, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (Oct) (2011) 2825–2830.
[66] H. Tamaki, A. Yoshinari, A. Okayama, et al., High Strength Ni-Base Superalloy for Directionally Solidified Castings, US Patent 6,051,083 (2000) 1-24.
[67] S.K. Naik, High Strength Nickel Base Single Crystal Alloys, US Patent, 4,885,216 (1989)1-14.
[68] T. Kobayashi, Y. Koizumi, T. Yokokawa, et al., Development of 4th generation SC superalloys without Re, J. Jpn. Inst. Metals 69 (2) (2005) 272–275.