From creep-life prediction to ultra-creep-resistant steel design
An uncertainty-informed machine learning approach

Wang, Chenchong; Wei, Xiaolu; van der Zwaag, Sybrand; Wang, Qiang; Xu, Wei

DOI
10.1016/j.actamat.2025.121073

Publication date
2025

Document Version
Final published version

Published in
Acta Materialia

Acta Materialia 292 (2025) 121073

Contents lists available at ScienceDirect

Acta Materialia

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

From creep-life prediction to ultra-creep-resistant steel design: An uncertainty-informed machine learning approach

Chenchong Wang [a], Xiaolu Wei [a,*], Sybrand van der Zwaag [b], Qiang Wang [c], Wei Xu [a]

[a] State key laboratory of rolling and automation, Northeastern University, Shenyang, Liaoning 110819, China
[b] Novel Aerospace Materials Group, Faculty of Aerospace Engineering, Delft University of Technology, 2629 HS, Delft, the Netherlands
[c] Key Laboratory of Electromagnetic Processing of Materials (Ministry of Education), Northeastern University, Shenyang, Liaoning, 110819, China

ARTICLE INFO

Keywords:
Creep life
Machine learning
Transfer learning
Alloy design
Uncertainty

ABSTRACT

In this research a machine learning model incorporating uncertainty to enhance the creep-life prediction and high-throughput design of creep-resistant steel is proposed. The framework integrates key physical metallurgical parameter linked to precipitate coarsening and applies transfer learning to correlate short-time tensile properties with the creep performance, all within a Bayesian convolutional neural network. Unlike conventional machine learning models, which often lack an assessment of prediction credibility, this uncertainty-based approach offers more accurate and stable predictions while also providing a measure of prediction credibility. By combining the model with a genetic algorithm, the framework achieves a balance between creep life optimization and uncertainty, thereby supporting robust alloy design. The validation on newly developed martensitic heat-resistant steels with tolerable prediction uncertainty showed excellent alignment between predicted and experimentally determined creep life, underscoring the effectiveness of the framework. These findings highlight the critical role of uncertainty modeling in advancing machine learning applications for alloy design.

1. Introduction

Creep resistance is a critical property of metallic components that must operate under long-term high-temperature conditions. In industries such as energy, aerospace, and power generation, metallic components are subjected to extreme heat and must maintain their structural integrity. For example, in thermal power plants, heat-resistant steel must continuously withstand elevated temperatures for hundreds of thousands of hours. Enhancing the creep life of these materials is a major design challenge. However, traditional creep testing methods are both time-consuming and costly, limiting the feasibility of discovering improved materials through experimental trial and error alone. Considering the growing demand for novel high-performance steels suitable for operation under extreme conditions, it is important to accurately predict creep life and design alloys with better endurance [1–4].

Over the years, various models have been developed to predict creep life [5–12], which are generally classified into two categories [13]: (1) empirical models based on endurance strength extrapolation and (2) models derived from physical mechanisms. Empirical methods, such as isothermal extrapolation and the time-temperature parameter, estimate the long-term creep life by extrapolating short-term test data. In particular, the Larson-Miller parameter [14] and theta method [15–17] are widely used because of their simplicity. However, these models require extensive experimental data and often lack general applicability across different materials. In addition, long-term microstructural changes and complex fracture mechanisms can reduce the accuracy of these predictions, particularly for extended service lives. Physical mechanism-based models, such as the creep power law, predict the creep life based on the relationship between the minimum creep rate, stress, and temperature, often incorporating the Monkman–Grant relationship [18]. Although these models offer a scientific basis, their accuracy depends on parameters such as the creep activation energy and stress exponent, which are dependent on stress and temperature, limiting their generalizability. Although these models have been improved by refinements, such as Wilshire's incorporation of tensile strength, they still cannot capture the full complexity of creep mechanisms [19–24].

Recent advances in physical metallurgy (PM) have underscored the importance of microstructural changes during creep, including grain-

* Corresponding author.
E-mail addresses: wangchenchong@ral.neu.edu.cn (C. Wang), weixl@smm.neu.edu.cn (X. Wei), S.vanderZwaag@tudelft.nl (S. van der Zwaag), wangq@mail.neu.edu.cn (Q. Wang), xuwei@ral.neu.edu.cn (W. Xu).

https://doi.org/10.1016/j.actamat.2025.121073
Received 8 January 2025; Received in revised form 18 April 2025; Accepted 21 April 2025
Available online 22 April 2025

boundary void formation, dislocation density changes, and precipitate coarsening, all of which significantly affect material performance. Some models, such as that proposed by Basoalto et al., incorporate microstructural evolution to simulate damage accumulation during creep based on material-specific characteristics [5]. However, these models often simplify the microstructural characterization and overlook the diverse effects of the alloy composition on creep properties, thereby limiting their value in practical alloy design. In addition, creep-damage models based on continuum damage mechanics have been actively developed over the past two decades to improve creep-life estimation [25–32]. Despite these advances, these models have limitations. Their practical applicability in alloy design is restricted because they may not adequately account for the specific damage mechanisms of new alloy compositions.

In recent years, computational and data sciences have developed machine learning (ML) as a promising tool for predicting material properties [33–35]. Unlike traditional models that rely on extensive data fitting on relatively small and focused datasets, ML extracts patterns from large, diverse datasets, thereby addressing the limitations of parameter-heavy models. For creep-life prediction, researchers have begun incorporating multiple features, such as composition, processing, and microstructural parameters [36–40]. For example, Liu et al. developed a machine learning model for nickel-based superalloys, clustering creep data and selecting optimal algorithms for different categories to improve prediction accuracy [36]. Wang et al. applied ML to Cr–Mo martensitic steel, enhancing the model accuracy by incorporating processed creep parameters [37]. Shin et al. applied thermodynamic calculations to extract PM information for alumina-forming austenitic stainless steel and performed a correlation analysis between the input features and Larson–Miller parameters using various algorithms, providing valuable insights into the integration of expert knowledge into creep prediction models [38]. However, despite these advancements, traditional ML models still face challenges in accurately capturing the complex interactions between composition, processing parameters, microstructure, and creep performance.

The rise of deep learning has led to increased research on the application of the aforementioned techniques to predict steel properties [41–45], including the creep life. For instance, Xiang et al. developed a deep-learning model for an HP40Nb heat-resistant alloy, combining a multilayer perceptron and convolutional neural network (CNN) to integrate numerical data and microstructural features [46]. This innovative approach has the potential to significantly enhance prediction accuracy. Similarly, Zhang et al. applied deep neural networks to predict the creep life of 316 austenitic stainless steel, further demonstrating the capabilities of deep learning in this field [47]. However, deep-learning models often face challenges, such as overfitting, particularly for small datasets. This is especially relevant for the creep performance, which is heavily influenced by the microstructural evolution. The limited availability of creep-life data complicates matters because deep-learning models typically require large datasets for effective performance. In addition, neural networks may struggle to generalize new unseen data, which increases the uncertainty of their predictions. To address these challenges, researchers have begun incorporating uncertainty into creep-life models. Yoo et al. introduced a Bayesian neural network model designed specifically to quantify the prediction uncertainty [48]. Meanwhile, Peng and Kwon et al. highlighted the importance of uncertainty quantification in assessing the reliability of ML models, particularly for complex properties such as creep [49,50]. By integrating uncertainty into these models, researchers can better evaluate their extrapolation capabilities for small data sets, providing a more robust foundation for design optimization.

In summary, predicting creep life poses significant challenges owing to the limited availability of data and the complexity of the underlying mechanisms. Traditional models, whether empirical or based on physical mechanisms, often struggle with accuracy and generalizability. Although ML and deep learning techniques show promise, their performance is limited by data scarcity. Future research should aim to create reliable models that perform well with small sample sizes while also incorporating uncertainty quantification and insights from microstructural data and PM to improve their interpretability and practical applications.

Based on this background, this study introduces a novel framework for predicting the creep life of martensitic heat-resistant steel and designing high-performance alloys by integrating the principles of PM with uncertainty quantification within a deep-learning framework. The framework combines PM parameters related to precipitate coarsening with transfer learning (TR), which connects short-time tensile properties with creep performance. A Bayesian convolutional neural network (BCNN) is used to provide accurate predictions along with quantified uncertainty, even when working with limited datasets. For alloy design, a genetic algorithm optimizes the composition and processing conditions to maximize the creep life while minimizing uncertainty. This PM-TR-BCNN framework facilitates efficient alloy design using restricted data.

2. Methods

2.1. Dataset and feature analysis

The dataset for this study was derived from creep data sheets of the NIMS structural materials data sheets [51,52]. This collection includes various high-Cr martensitic heat-resistant steels such as 9Cr-1Mo, 9Cr-1Mo-V-Nb, 9Cr-0.5Mo-1.8W-V-Nb-B, and 12Cr-2W-0.4Mo-1-Cu-Nb-V. The dataset comprises 704 samples of heat-resistant steel, each with 23 input features and one output feature. The input features include 16 composition parameters and five processing parameters—normalizing temperature (NT), normalizing time (Nt), normalizing cooling rate (NCR), tempering temperature (TT), and tempering time (Tt)—and two test conditions—test stress (TestS) and test temperature (TestT). The output feature, creep life, is expressed in logarithmic form to reduce scale variations, linearize its relationship with stress and temperature, and lessen the influence of extreme values. This transformation also preserves relative differences and reflects the physical nature of time-dependent failure in steels. Table S1 summarizes the input and output features, along with their respective ranges. In addition, a high-temperature short-time tensile property dataset corresponding to these steels was collected [51,52], which contained 243 samples. Each sample also had composition, processing parameters, and TestT. The output features for this dataset are the yield strength (YS), ultimate tensile strength (UTS), elongation (EL), and area reduction (RA), the details of which are presented in Table S2. For samples not having received a tempering treatment, the tempering temperature and the tempering time both were set to zero. Finally, z-score standardization was applied to normalize the data.

Given the complexity of the relationship between composition, processing parameters, test conditions, and creep life in heat-resistant steels, both the pearson correlation coefficient (PCC) and mean decrease accuracy (MDA) [33] were used to assess the impact of each feature on creep life, as illustrated in Fig. 1. The results of the PCC and MDA analyses both indicated that the TestT and TestS were the most influential factors on creep life, each showing a negative correlation. The MDA values for these two parameters were above 90 %, indicating their strong predictive relevance. Apart from the test conditions, the next most influential features were the concentrations of elements associated with the primary precipitation phase, such as N, V, and Nb, and the heat treatment conditions, such as tempering temperature. These feature importance results align well with the established metallurgical knowledge of heat-resistant steels, thereby supporting the high quality of this dataset. While most features exhibit low PCC values, this is expected because PCC captures only linear dependencies. However, the relationships between alloy composition, processing conditions, and creep life are often nonlinear and involve complex interactions. To

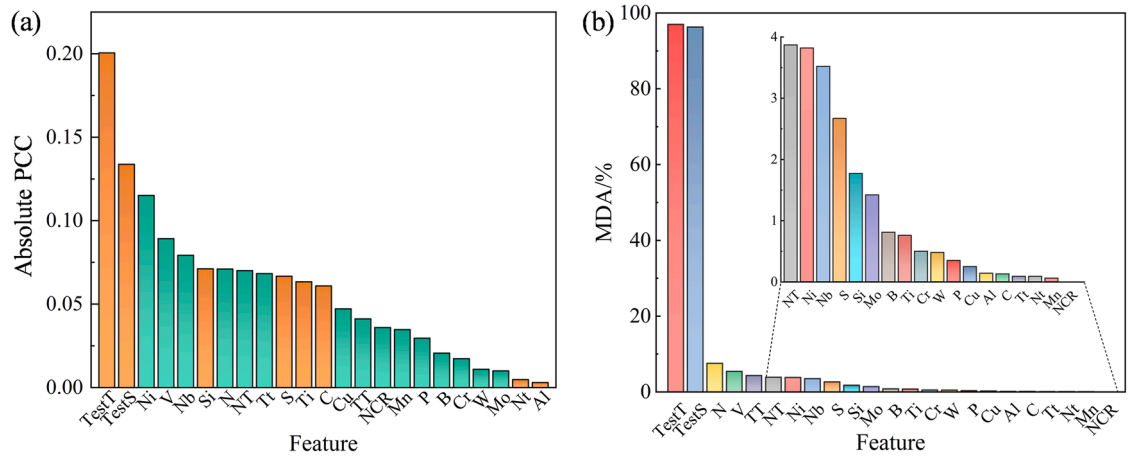

Fig. 1. (a) Absolute PCC and (b) MDA importance assessment of all features in the creep dataset. Orange and green bars in (a) represent features with negative and positive correlations with creep life, respectively.

account for these effects, maximal information coefficient (MIC) analysis was applied, confirming that all features contribute to creep life prediction (Fig. S1). Furthermore, the MDA analysis supports that all features contribute positively to model performance. These findings justify the inclusion of all features in the model without additional feature selection. A more detailed discussion is provided in Supplementary 2. In addition, potential redundancy and collinearity were evaluated by calculating PCC among input features, which revealed several highly correlated pairs (e.g., Nb–N, V–N, V–NT, Nb–NT), as shown in Fig. S2. Although principal component analysis (PCA) is a common processing strategy, initial tests showed that dimensionality reduction did not significantly enhance model performance. Retaining the original features preserved both physical interpretability and robust uncertainty calibration, so PCA was not adopted.

2.2. Physical metallurgical information mining

The heat-treatment process for martensitic heat-resistant steel typically involves normalization followed by high-temperature tempering. The normalization stage creates a martensitic microstructure that is subsequently tempered. This approach was developed to enhance the toughness and thermal stability of steel for long-term high-temperature applications. During tempering, the dislocation density in the martensitic matrix decreases and secondary phases precipitate from the matrix. In 9 %–12 % Cr martensitic heat-resistant steel, the primary precipitation phases after tempering are MX carbonitrides and $M_{23}C_6$ carbides. MX carbonitrides, generally tens of nanometers in size, precipitate within the martensitic laths, whereas $M_{23}C_6$ carbides, approximately 100 nm in size, form primarily along the interfaces. The strengthening of martensitic heat-resistant steel relies predominantly on solid-solution and precipitation strengthening. Consequently, the PM information relevant to these mechanisms is essential for ML models that focus on creep-life prediction and alloy design.

For solid-solution strengthening modeling, an implicit approach can be adopted using a transfer-learning model that correlates short-time tensile properties with long-term creep behavior. The short-time tensile properties provide insights into the differences in solid-solution and initial precipitation strengthening properties observed across various compositions before high-temperature service. Over time, as these steels are exposed to service conditions, the effectiveness of precipitation strengthening diminishes. The MX and $M_{23}C_6$ phases coarsen, decreasing their contribution to strengthening. Additionally, during service, the Laves phase precipitates and grows within the heat-resistant steel matrix. Initially, these precipitates are fine and provide some precipitation strengthening. However, they coarsen much more rapidly than the MX and $M_{23}C_6$ phases, significantly reducing their strengthening contribution. Additionally, as the Laves phase forms, Mo and W are depleted from the matrix, weakening the solid-solution strengthening effect. As the Laves phase generally appears during the later stages of creep, this study primarily focused on calculating the strengthening contributions of the MX and $M_{23}C_6$ phases.

The evolution of the strengthening contributions of the MX and $M_{23}C_6$ phases was quantified by the precipitation strengthening factor described by Lu et al. [4], which accounts for changes in precipitate size, density, and distribution. The strength contribution of precipitation strengthening is expressed as [53]:

$$\sigma_p = \frac{\alpha G b}{L} \tag{1}$$

where α is a constant, G is the shear modulus, b is the Burgers vector, and L is the average particle spacing. Given the narrow range of steel compositions in the dataset used in this study, G and b were assumed to be composition-independent. Thus, σ_p is inversely proportional to L, where $1/L$ denotes the strengthening effect.

$$\sigma_p \propto \frac{1}{L} \tag{2}$$

Over long-term service, the particle spacing changes are driven mainly by precipitate coarsening. Therefore, the coarsening kinetics was used to evaluate the change in particle spacing [4,53]:

$$L = \frac{r}{\sqrt{f_p}} = \frac{\sqrt[3]{r_0^3 + Kt}}{\sqrt{f_p}} \tag{3}$$

$$r_0 = \frac{2\gamma}{\Delta G_v} \tag{4}$$

where r is the average particle radius, f_p is the equilibrium volume fraction of the precipitate at service temperature, r_0 is the critical nucleation size, γ is the matrix-precipitate interface energy, ΔG_v is the volume thermodynamic driving force for precipitation, K is the coarsening rate, and t is the time held at high temperature. The parameters γ, f_p, ΔG_v, and K were determined using the TCFE9 and MOBFE4 databases in Thermo-Calc software, while t was taken as the creep life of the heat-resistant steel. Substituting these values into the Eqs. (3)-(4) yielded 1/L. The $1/L$ parameters for the two precipitated phases were combined and processed numerically by scaling by 10^6 and applying a natural logarithm to yield the final precipitation strengthening factor, named PF. Fig. 2 illustrates the distributions of TestS, TestT, and the calculated PF and creep life across the dataset. The PF followed a normal distribution, showing an inverse correlation with temperature and a direct

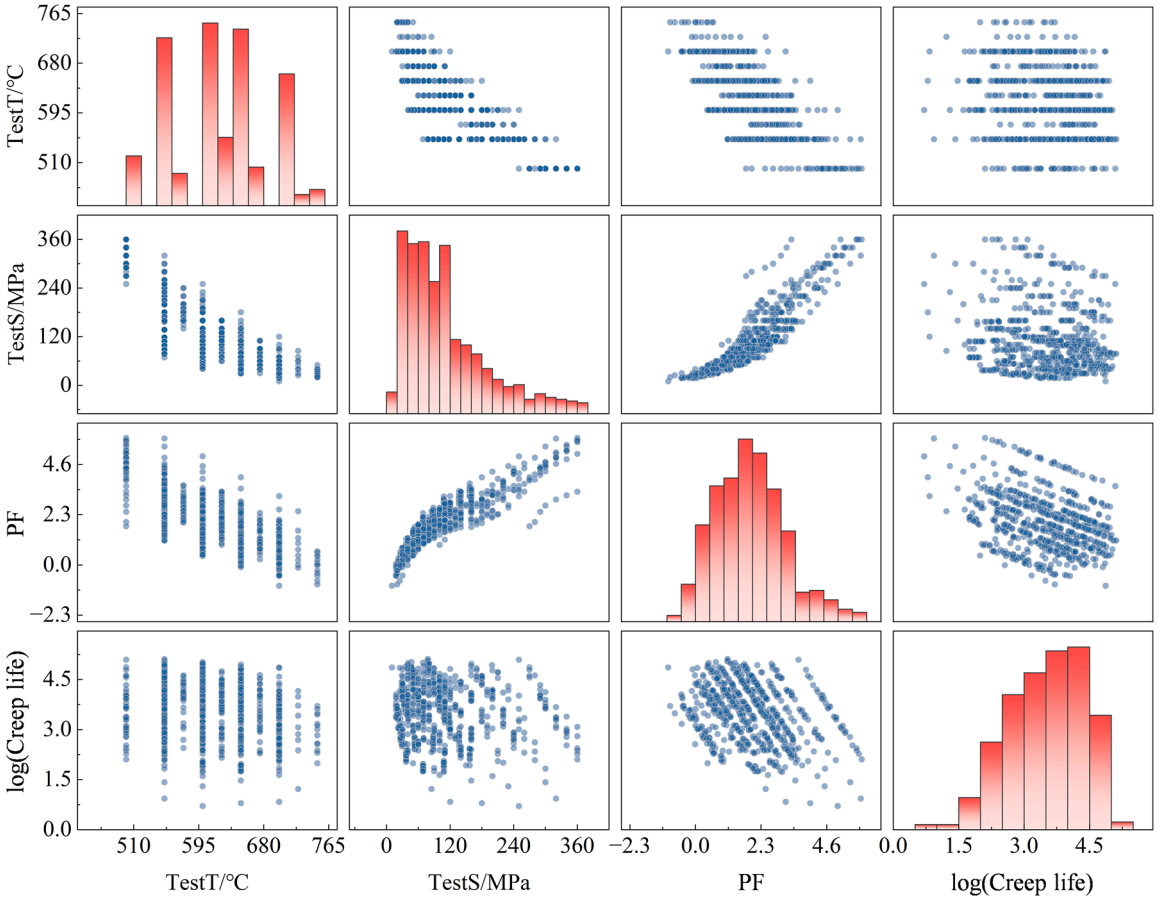

Fig. 2. Distribution of TestT, TestS, calculated PF and creep life.

correlation with stress. To further assess the role of PF in predicting creep life, a detailed correlation analysis was conducted. PCC value between PF and log(creep life) was found to be 0.327, the highest among all input features. This strong relationship underscores PF's significance, as it integrates key metallurgical parameters such as composition, heat treatment conditions, and time-dependent precipitate coarsening kinetics, which directly influence the degradation of precipitation strengthening during creep. In addition, MIC and MDA analyses were applied. The MIC score for PF was 0.227, the highest among all features, indicating a substantial nonlinear association with creep life. While the MDA value was 96.9 % (Fig. S3), even exceeding that of TestT and TestS. These analyses highlight PF's crucial role in connecting microstructural evolution to creep performance. In addition, to assess the contributions of the parameters that form PF, two analyses were performed. First, the relative impacts of these parameters on predicted creep life were evaluated using PCC and MIC. Next, an alternative model was built where the PF sub-components (γ, f_p, ΔG_v, and K) were introduced as separate variables instead of a single composite parameter. The analysis in Supplementary 4 shows that combining them into PF is a more effective approach.

2.3. PM-TR-BCNN framework

An uncertainty prediction framework integrating PM and TR was developed to address the complex relationships among composition, processing parameters, test conditions, and creep performance. In this framework, a TR module leverages the correlation between the short-time tensile properties and creep life, while the PM module calculates PF, as illustrated in Fig. 3(a). Additionally, a direct module links the composition, processing variables and test conditions to the creep life. These three modules were implemented within a BCNN to form the PM-TR-BCNN creep-life uncertainty prediction framework. BCNNs enhance traditional CNNs by integrating Bayesian inference, which treats network weights as probability distributions rather than as fixed values. This probabilistic framework enables the BCNN to effectively manage uncertainty and improve generalization, particularly for complex datasets. During training, the BCNN utilizes variational inference to estimate the parameter uncertainty, resulting in multiple distributions that yield a range of prediction outputs. Consequently, the BCNN provides quantifiable uncertainty measures for each prediction, which are crucial for assessing model reliability.

The implementation of the PM-TR-BCNN framework proceeds as follows. (1) Using a high-temperature, short-time tensile dataset, a BCNN source model is first constructed to link the steel composition, processing, and TestT with tensile properties. (2) The feature extraction layers from the source model are transferred to the PM-TR-BCNN as a fixed "transfer feature layer," which remains frozen in subsequent training steps and is connected to an adjustable fully connected layer, forming the TR module with composition, processing, and TestT as inputs. (3) The PM module accepts the composition, processing, and TestT/TestS inputs to predict the PF. (4) The Creep module uses the composition, processing, and TestT/TestS inputs and outputs the results through an adjustable, fully connected layer. (5) Finally, the outputs from the TR, PM, and Creep modules are combined to predict the creep life, thus forming a complete PM-TR-BCNN architecture. During training, the loss function of the model includes terms for both the PF and creep life, as shown in Fig. 3(a).

The model parameter settings are detailed in Fig. 3(b). For the source model, the composition, processing, and TestT data from a short-time tensile dataset (243 samples, 22 dimensions) were reshaped into a 5 ×

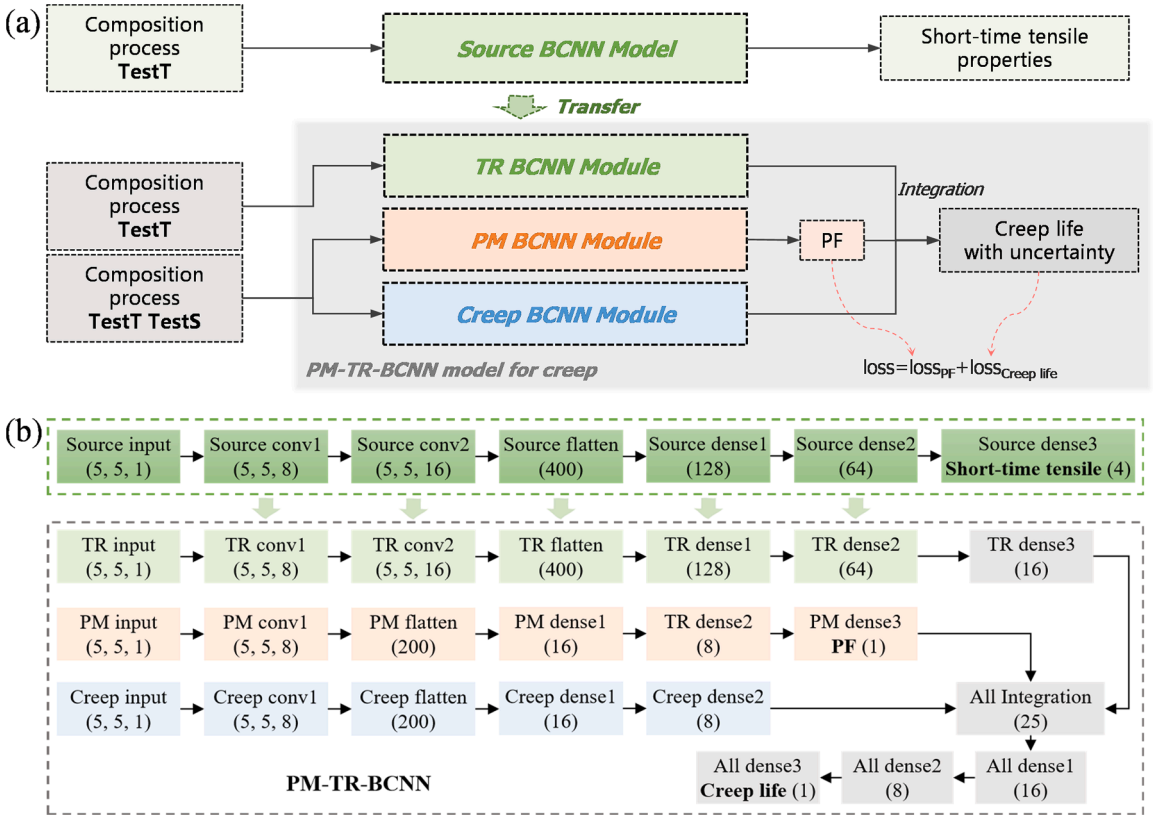

Fig. 3. (a) Schematic of the creep life uncertainty prediction framework combining physical metallurgical knowledge and transfer learning (PM-TR-BCNN). (b) Network architecture of PM-TR-BCNN model.

5 matrix as the BCNN input. The model outputs four tensile properties: YS, UTS, EL, and RA. The TR module uses the same input format as the source model, whereas the PM and Creep modules use 23-dimensional creep data (composition, processing, TestT, and TestT inputs) reshaped into a 5 × 5 matrix.

For the short-time tensile dataset, the samples were divided into training and testing sets in a 9:1 ratio. This process was repeated ten times randomly, and the optimal model was selected to establish the transfer module of the PM-TR-BCNN. For the creep dataset, 140 samples were randomly chosen as the validation set from a total of 704 samples. The remaining samples were then divided into training and testing sets in an 8:2 ratio. This data-splitting process was repeated randomly 50 times.

During model training, 5000 iterations were used with the mean squared error (MSE) loss function, a learning rate of 10^{-4}, and the Adam optimizer. These hyperparameters were determined through multiple rounds of model tuning to ensure stable convergence without overfitting. The model architecture was also carefully refined to enhance generalizability. As shown in Fig. S6, for a representative random split, the training and validation loss curves converge smoothly, with no signs of divergence. These strategies were consistently applied across all 50 random splits, demonstrating that the model is robust and effectively avoids overfitting. For the BCNN, a standard multivariate normal distribution with zero mean and unit variance was used as the prior for the weight distributions, following the default setting in the TensorFlow Probability framework. Once model training was complete, predictions were made 100 times for the same composition, processing, temperature, and stress conditions. The final predicted creep life was calculated as the mean of 100 predictions, with the standard deviation representing the uncertainty of the prediction.

For comparison, the corresponding PM-TR-CNN model (using a CNN as the base model instead of BCNN) were trained as reference models following a construction methodology similar to that used for PM-TR-BCNN. In addition, PM-TR-BCNN was compared with common machine learning methods such as random forest. While these models achieved comparable accuracy in some cases, they lack built-in uncertainty quantification (Fig. S5). Bayesian inference adds computational complexity due to multiple forward passes for uncertainty estimation. Compared to non-Bayesian models, the PM-TR-BCNN model requires more training and inference time. Despite this, this extra effort is justified by the benefit of reliable uncertainty estimates. Meanwhile, the PM-TR-BCNN model was compared with an alternative uncertainty quantification strategy of ensemble modeling, demonstrating that the benefit of introducing uncertainty is not limited to PM-TR-BCNN. While the PM-TR-BCNN model maintained good uncertainty calibration, both methods significantly outperformed deterministic models. All modeling, data preprocessing, and model training were conducted using Python, TensorFlow, TensorFlow Probability, and Scikit-learn. To evaluate the model-prediction performance, two key metrics were used: the squared correlation coefficient (R^2) and mean absolute error (MAE), which were calculated as follows:

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |f(x_i) - y_i| \quad (6)$$

2.4. Alloy design and validation

After establishing the creep-life prediction model, the steel was redesigned using a multi-objective genetic algorithm (NSGA II) to optimize the alloy solution at 650 °C and 140 MPa. 650 °C/140 MPa was selected as the target condition based on its industrial relevance, data

availability, and suitability for model benchmarking. This temperature is representative of service environments in advanced ultra-supercritical power plants. Although actual operating stresses are often lower, testing at 140 MPa allows accelerated creep evaluation within practical timescales and is consistent with commonly reported experimental conditions. Furthermore, a substantial number of creep tests in the dataset are conducted under this condition, enabling statistically robust performance comparison with existing alloys. Thus, 650 °C/140 MPa provides a realistic and reliable basis for validating the proposed alloy design approach. Further description is provided in Supplementary 6.

To meet the aims of this study, the optimization objectives of NSGA II include maximizing the creep life while minimizing the prediction uncertainty. During the design process, considering that the increase of solid solution strengthening elements Mo and W will increase the risk of Laves phase precipitation, the Mo equivalent ($Mo_{eq} = w[Mo] + 0.5w[W]$) is limited to less than 1.9 %. This threshold was determined by thermodynamic calculations. The genetic algorithm was parameterized with a population size of 20 individuals, 1000 iterations, a mutation rate of 0.1, and a crossover rate of 0.95. The algorithm was executed ten times to prevent the local optima from influencing the final results. To ensure the reliability and effectiveness of the designs generated by the prediction model, only models with an R^2 greater than 95 % for both the training and testing sets were utilized as objective functions in the NSGA II.

In addition, factors related to low-concentration impurity elements must be considered during the design process. As shown in Fig. 1, both PCC and MDA analyses indicate that sulfur (S) and phosphorus (P), present in low concentrations, have a limited impact on creep life.

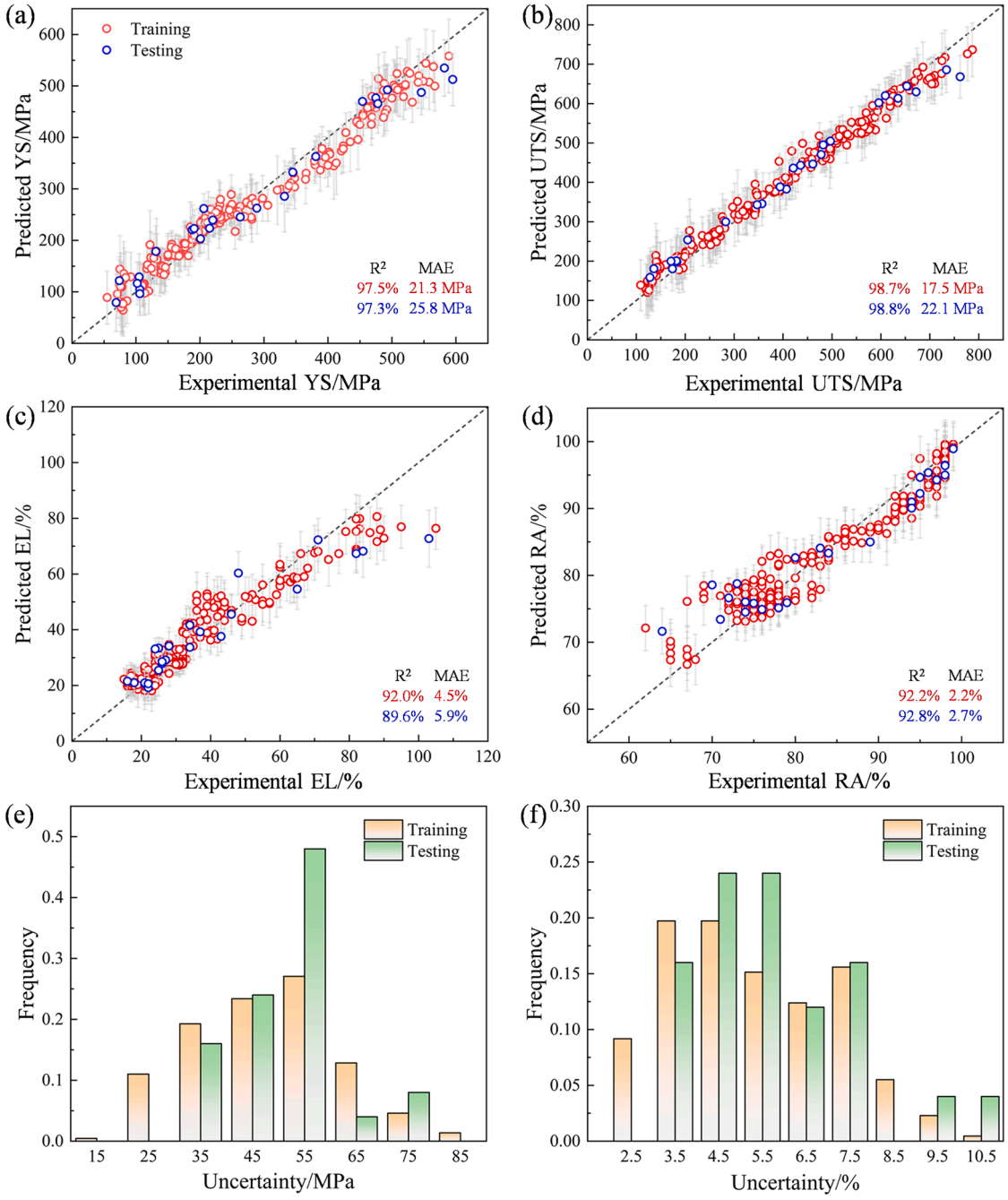

Fig. 4. Short-time tensile properties predicted by the BCNN tensile model compared with the experimental results: (a) YS, (b) UTS, (c) EL, and (d) RA. Uncertainty distribution for (e) UTS and (f) EL prediction.

Therefore, in the design process, the contents of S and P were fixed at their dataset mean values, and alloy solutions resulting in elevated levels of these elements were excluded.

Finally, the NSGA II and PM-TR-BCNN models were combined to design three new steels with creep life exceeding 288.5 h under the specified conditions of 650 °C and 140 MPa. These alloys were manufactured, and creep tests were conducted on appropriately treated samples. The three designed alloys were smelted and cast in a vacuum induction furnace to produce 50 kg ingots, which were then forged at 1100 °C into billets with a cross-sectional size of 200 × 135 mm^2. Subsequently, the billets were heated at 1200 °C for 4 h, hot rolled to a thickness of 20 mm, and then subjected to laminar cooling. Subsequent heat treatments, including normalization and tempering, were performed according to the temperature and time specified by the design solution. The creep tests were performed using samples with dimensions consistent with those reported in the NIMS creep datasheet.

The microstructures of three designed alloys were characterized using scanning electron microscopy (SEM) and transmission electron microscopy (TEM). For SEM analysis, the samples were metallographically prepared and etched using a solution containing 2.5 g of ferric chloride, 15 ml of hydrochloric acid, and 100 ml of water. Observations were conducted on a JEOL JSM 7200F microscope with a field emission gun. For TEM analysis, samples were prepared through twin-jet electropolishing in a solution of 10 % perchloric acid and 90 % alcohol at approximately −20 °C. The observations were carried out using an FEI Talos F200X microscope equipped with an energy-dispersive spectroscopy (EDS) system for compositional analysis.

3. Results

3.1. Creep life prediction

Initially, a BCNN source model was developed to predict short-time tensile properties, as detailed in Section 2.3. Fig. 4(a-d) illustrate the prediction results of the optimal model, demonstrating accurate predictions for all four tensile properties. Notably, the EL prediction performance is slightly lower than those of the other three metrics (with R^2 values exceeding 90 %). The testing set accuracy for the UTS is impressive, surpassing 98 %, with an MAE of only 22 MPa. The average prediction outcomes of the BCNN model across the data partitions are summarized in Table 1, highlighting its robust predictive capabilities for modeling the short-time tensile properties of heat-resistant steel. In addition to its high predictive accuracy, the BCNN model offers a significant advantage over traditional statistical learning methods and CNNs by providing an uncertainty estimate for each prediction (as shown by the error bars in Fig. 4(a-d)). Fig. 4(e) and 4(f) further present the uncertainty distribution of the source model predictions for the UTS and EL, which appear to follow a roughly normal distribution. The uncertainty profiles for both the training and testing sets exhibit a high degree of similarity, indicating consistency in model performance. Most of the uncertainty values were below 55 MPa for UTS and 7.5 % for EL, demonstrating the reliability of the model and its ability to quantify the prediction confidence effectively.

After successfully developing a precise source model to predict tensile properties, it was adapted to create a PM-TR-BCNN model. During model training, PF and creep life were included as objectives, resulting in outputs that include these two critical parameters. Fig. 5 illustrates the predicted results for PF and creep life derived from the optimal PM-TR-BCNN model across 50 random splits. Fig. 5(b) compares the experimental and predicted creep life (log(Creep life)) for different TestT (500–750 °C) and TestS (10–360 MPa). It can be seen that for the creep life, the predictions closely correspond to the experimental values. As a result, the model achieved an R^2 value of 96.5 % and a MAE of 0.127 for the creep-life validation set. The PF predictions are similarly robust, achieving an R^2 value of 94.7 % for the validation set. Fig. 5(c) further illustrates the uncertainty distribution of the creep-life predictions generated using this model. Most of the prediction uncertainties across the training, testing, and validation sets remained below 0.3, with most values concentrated within the range of 0.15–0.25. Typical uncertainties are ±0.2 log(Creep life) across the entire creep-life range. This indicates that the PM-TR-BCNN model delivered accurate predictions for the various samples, along with a specific prediction confidence value for each sample. Table 2 summarizes the average prediction results from the model across 50 random splits, revealing an average R^2 for the validation set of above 95 % with minimal standard deviations. The model was also evaluated across different stress and temperature ranges. While it generally performed well, accuracy decreased slightly under extreme conditions, particularly at high temperature (>725 °C) or very low/high stresses (<50 MPa or >250 MPa), as shown in Fig. S7 and S8. This was accompanied by higher uncertainties, likely due to limited data for these conditions. However, the PM-TR-BCNN model effectively flags these cases, indicating reduced reliability and guiding cautious interpretation of predictions at the extremes. Overall, these findings underscore the reliability of the PM-TR-BCNN framework and demonstrate its capacity to provide stable and accurate predictions while quantifying the associated uncertainty, which is essential for understanding the model's predictive reliability for practical applications.

From the large database involving multiple test temperatures, we now focus on the subset of data for a TestT of 650 °C (142 samples, covering seven steel grades). Fig. 5(d) compares the experimental and predicted creep life values versus test stress for the three best-performing heat-resistant steels at 650 °C, named alloys E1–E3 here (the lowest performing grades in the subset have less than 50 % of the load bearing capacity of these three grades). Taking these three heat-resistant steels as examples, their creep lives gradually decrease with increasing stress. The model accurately captured this trend, and the predicted creep lives are in good agreement with the experimental values. The prediction uncertainties were distributed between 0.15 and 0.32. Further alloy optimization design was performed at typical test conditions of 650 °C and 140 MPa.

Table 1
Mean prediction results from the BCNN source models.

Tensile properties	Training set		Testing set	
	R^2/ %	MAE/MPa, %	R^2/ %	MAE/MPa, %
YS (MPa)	97.4 ± 0.2	21.1 ± 0.7	96.8 ± 1.4	22.6 ± 3.3
UTS (MPa)	98.8 ± 0.1	16.8 ± 1.0	98.2 ± 1.0	19.8 ± 3.9
EL (%)	92.5 ± 0.6	4.3 ± 0.2	91.2 ± 4.2	4.7 ± 0.9
RA (%)	92.4 ± 0.2	2.2 ± 0.05	90.5 ± 3.2	2.6 ± 0.4

3.2. Alloy design considering uncertainty

As highlighted in the previous section, the predicted results demonstrate the capability of the PM-TR-BCNN model to capture key trends in the creep-life performance. This section discusses the application of the optimal PM-TR-BCNN model to alloy design. Although the analysis focuses on creep performance optimization at 650 °C and 140 MPa, the methodology is broadly applicable to other high-temperature and stress scenarios, provided that appropriate input data and constraints are used. The strategy for alloy design involved searching within the compositional and processing parameter ranges available in the database for alloy solutions that exhibit optimized predicted creep lives with sufficiently low scatter (uncertainty) based on the NSGA-II algorithm.

Before discussing the design of new compositions, the experimental creep performance and model (back-) predictions for the best-performing alloys (E1–E3) at 650 °C and 140 MPa are briefly reviewed, as shown in Fig. S9. The model slightly overestimated the creep life of alloys E1–E3, with the best-performing alloy (E1) having a predicted life of $10^{2.58}$ h. The prediction uncertainties for these alloys range from 0.24 to 0.29. These uncertainties, which reflect the model's

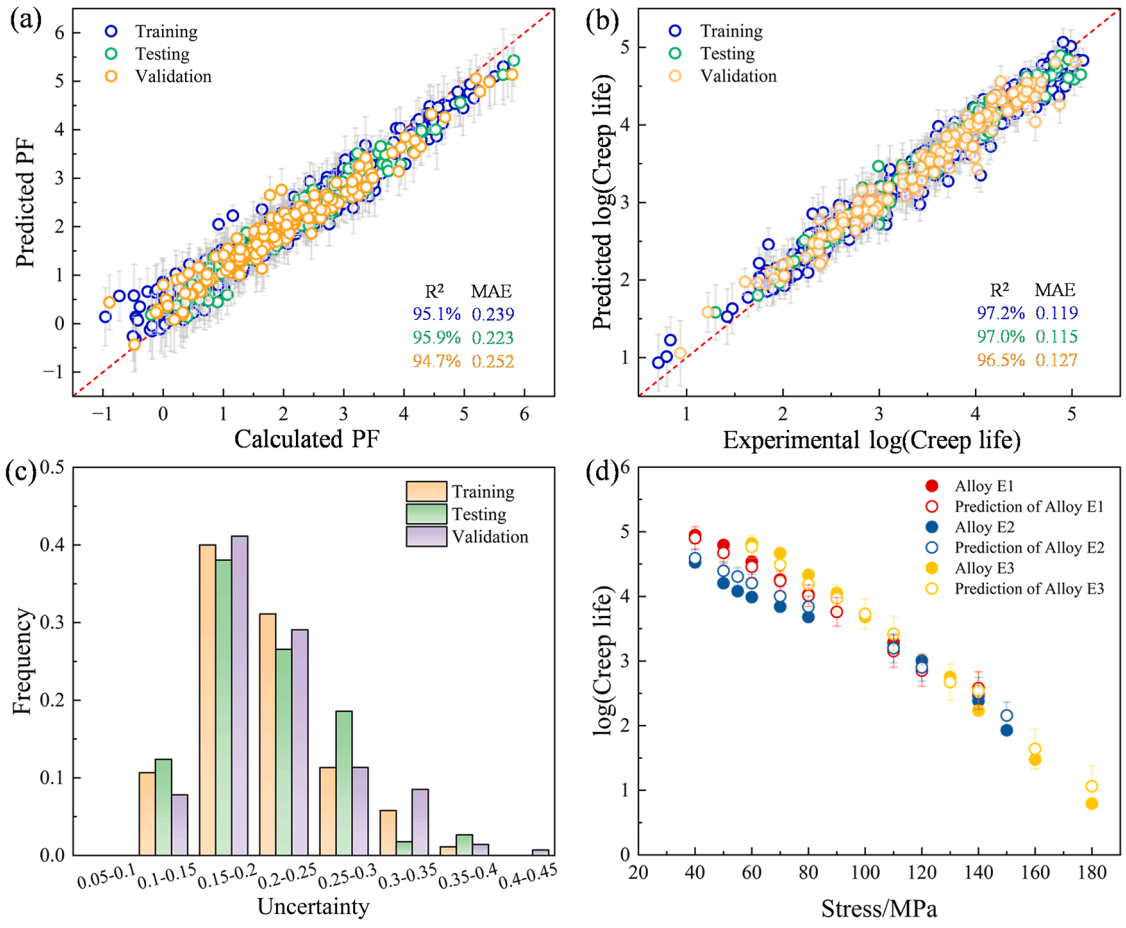

Fig. 5. Predicted results by the PM-TR-BCNN model. (a) Precipitation strengthening factor PF. (b) log(Creep life). (c) Uncertainty distribution for creep life prediction. (d) Creep life at 650 °C.

Table 2
Mean prediction results from the PM-TR-BCNN models.

Creep	Training set		Testing set		Validation set	
	R^2/ %	MAE	R^2/ %	MAE	R^2/ %	MAE
log(Creep life)	97.3 ± 0.3	0.116 ± 0.005	95.3 ± 1.1	0.148 ± 0.015	95.5 ± 1.0	0.145 ± 0.015
PF	94.4 ± 1.3	0.236 ± 0.024	93.6 ± 2.0	0.248 ± 0.032	93.8 ± 1.7	0.248 ± 0.034

confidence in its predictions, are within an acceptable range for practical alloy design. Therefore, the design strategy used these predictions and uncertainties as benchmarks for improvement.

This design strategy generated a series of alloy solutions with a range of predicted creep lives and prediction uncertainties (blue points in Fig. 6), illustrating the tradeoff between creep life and uncertainty. The vast majority of designed alloys achieved predicted creep lives exceeding those of the existing alloys in the dataset (i.e., E1–E3), and showing a balance between the predicted life ($10^{2.61}$–$10^{3.34}$ h) and uncertainty (0.22–0.41). To identify the optimal solutions, three alloys (D1, D2, and D3) were selected from the dual objective set based on specific criteria. Specifically, D1 was chosen for its lowest prediction uncertainty (0.22), ensuring the highest prediction reliability and moderate creep life improvement ($10^{2.61}$ h). Alloy D2 shows a higher predicted creep life ($10^{2.75}$ h) with tolerable uncertainty (0.27), similar to that of the existing alloys E1–E3 (0.24–0.29). Alloy D3 shows the highest predicted life ($10^{3.34}$ h) but a higher uncertainty (0.41), i.e., the highest prediction scatter and lowest prediction reliability among the three alloys.

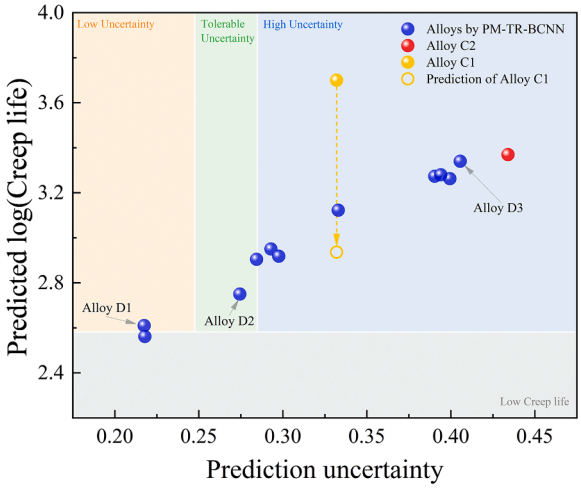

Fig. 6. Distribution of predicted creep life and uncertainty obtained by single-objective and dual-objective designs.

Additionally, for comparison, single-objective optimizations focusing solely on maximizing creep life were conducted using the PM-TR-CNN model (without prediction uncertainty considerations) and the PM-TR-BCNN model, resulting in alloys C1 and C2, respectively. The prediction performance of the PM-TR-CNN model are shown in Table S3. The predicted creep life of C1 ($10^{3.7}$ h, approximately 5000 h) was

deemed excessively high as it significantly deviates from observed trends in the 650 °C and 140 MPa dataset (maximum of 288.5 h). When evaluated using the PM-TR-BCNN model, the predicted creep life of this solution was reduced to $10^{2.94}$ h with a prediction uncertainty of 0.33, highlighting the risks of ignoring the prediction uncertainty in modeling. For alloy C2, the introduction of uncertainty modeling led to a more reasonable predicted life ($10^{3.47}$ h). However, because uncertainty was not included as an optimization criterion, this alloy exhibited a high prediction uncertainty (0.43). This behavior closely resembles that of alloy D3, demonstrating the risks associated with ignoring the prediction reliability of the alloy design.

Table 3 provides a detailed comparison of the compositions and processing parameters of the designed alloys, including those of the three existing E1–E3 alloys with long creep lives. The primary compositional differences among alloys D1–D3 lie in the contents of Cr, Mo, Cu, N, and W. For alloys D1–D3, the Cr contents are in the range of 9.5–11 wt. %, the W contents are consistently above 1.8 wt. %, and the Mo contents are between 0.57 and 0.9 wt. %. Compared to the alloys E1–E3, the W contents of D1–D3 are similar, but the Mo contents are higher, yielding Mo equivalents between 1.485 and 1.835. The Mo equivalents for the single-objective designs of alloys C1 and C2 are 1.695 and 1.545, respectively. Research in the field of heat-resistant steel, particularly on 9–12Cr alloys, has established that an optimal Mo equivalent of approximately 1.6 % results in peak creep strength. Additionally, for a given Mo equivalent, a higher W/Mo ratio is associated with an improved creep strength [54]. Although Alloys C1 and C2 have high Mo equivalents, they contain significantly low W contents, which degrades their overall creep performance. Therefore, the dual-objective design approach, incorporating uncertainty, improves the reliability of alloy designs.

3.3. Alloy validation

Alloys D1-D3 were experimentally tested to evaluate their creep performance under conditions of 650 °C and 140 MPa. First, the experimental tensile properties of alloys D1-D3 at 650 °C are in good agreement with the predicted values (Fig. S10). As for creep, each alloy was designed with different predicted lives and levels of prediction uncertainty. Two samples of each alloy were tested. Fig. 7 compares the experimental creep life of these alloys with the data for alloys E1–E3. Notably, the newly designed alloys D1 and D2 exhibited longer creep lives (i.e., better performance) than alloys E1–E3. Alloy D1 exhibits the lowest predicted uncertainty, with experimental creep lives of 334 and 575 h (log(creep life) of 2.52 and 2.76, respectively), which closely aligns with its predicted life of 407 h (log(creep life) of 2.61±0.22). This strong agreement suggests that low uncertainty in the design leads to predictable performance when the steel is manufactured and tested on a larger scale. Alloy D2, selected for its balance between a moderately long predicted life with moderate uncertainty (similar to those of alloys E1–E3), showed experimental creep lives of 540 and 616 h (log(creep life) of 2.73 and 2.79, respectively), which matches its predicted life of 562 h (log(creep life) of 2.75±0.27). This demonstrates that alloy designs with low uncertainty and a moderately long lifetime can achieve reliable performance. In contrast, the two samples of alloy D3, with the highest predicted creep life and uncertainty (indicating lower prediction confidence), exhibited creep lives of 135 and 662 h (log(creep life) of 2.13 and 2.82, respectively). This large experimental variation, along with the significant deviation from its predicted creep life of 2189 h (log(creep life) of 3.34±0.41), highlights the risks of designs that do not account for prediction uncertainty. Overall, alloy D2, selected based on its low uncertainty and predicted performance, demonstrated the best creep resistance, as confirmed by the creep curve data in Fig. S11.

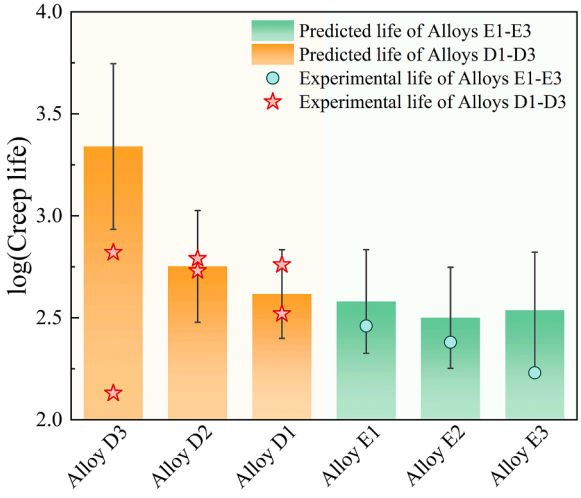

Fig. 7. Creep-life data (both experimental and predicted) for Alloys D1–D3 and Alloys E1–E3.

Microstructural analyses of alloys D1–D3 were conducted to further evaluate the observed performance differences. All alloys initially had tempered martensitic matrixes with various amounts of δ-ferrite, which is known to negatively impact creep resistance. The δ-ferrite contents of alloys D1, D2, and D3 were 2.3 %, 3.1 %, and 21.7 %, respectively (Fig. 8 (a, c, e)). This correlates with the calculated Cr equivalent values (11.3 wt. %, 11 wt. %, and 12.6 wt. %) and the thermodynamic results (Fig. S12(a)), suggesting that the large δ-ferrite content in alloy D3 contributes to its unstable creep performance. Additionally, all three alloys showed a tendency for Laves phase precipitation during testing (Figs. 8 and S12(b)), although this was not included in the PF calculations. The Laves phase can coarsen rapidly and deplete Mo and W, thereby weakening the alloy. The thermodynamic calculations confirm that alloy D3 has the highest volume fraction of Laves phase (Fig. S12 (b)), and its presence was also confirmed using TEM (Figs. 8(h) and (i)). In contrast, alloy D2 exhibited the lowest volume fraction of Laves phase. In summary, alloy D2, which contains only a small amount of δ-ferrite and exhibits the lowest tendency for Laves phase precipitation, shows the most stable creep performance.

Table 3
Composition and main processing parameters of newly designed alloys compared to those of existing alloys from the original dataset. Elemental fractions are in weight percentages, and temperatures are in Celsius.

Alloy	C	Si	Mn	Ni	Cr	Mo	Cu	V	N	Nb	W	B	NT	TT	Predicted (Experimental) log(Creep life)
D1	0.05	0.36	0.55	0.36	9.5	0.77	0.5	0.2	0.04	0.08	1.83	0.003	1050	790	2.61±0.22
D2	0.05	0.36	0.55	0.36	10	0.57	0.57	0.22	0.06	0.08	1.83	0.003	1050	790	2.75±0.27
D3	0.05	0.3	0.53	0.36	11	0.9	0.63	0.21	0.06	0.08	1.87	0.003	1100	790	3.34±0.41
C1	0.05	0.32	0.35	0.36	10.3	1.2	0	0.07	0.06	0.08	0.99	0.003	1100	800	3.70
C2	0.05	0.25	0.42	0.36	8.6	1.36	0.88	0.22	0.016	0.08	0.37	0.003	1100	752	3.48
E1	0.13	0.31	0.6	0.36	10.65	0.33	0.86	0.19	0.057	0.05	1.87	0.0024	1050	770	(2.46)
E2	0.11	0.27	0.59	0.33	12.1	0.34	0.82	0.19	0.066	0.06	1.82	0	1050	790	(2.38)
E3	0.098	0.29	0.42	0.13	9.5	0.36	0	0.19	0.046	0.062	1.74	0.0024	1100	780	(2.23)

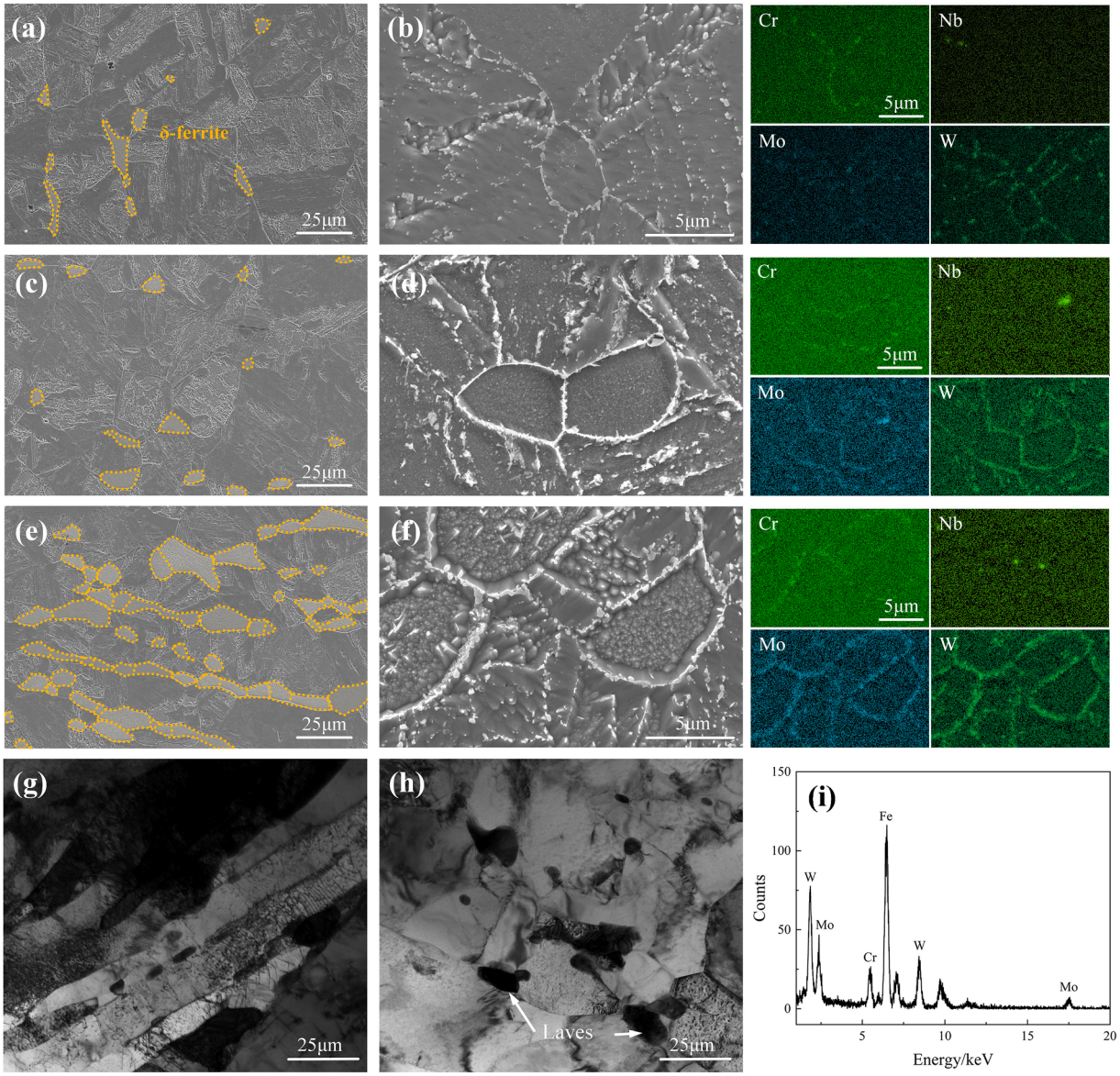

Fig. 8. SEM micrographs of (a) as-tempered Alloy D1, (b) Alloy D1 after creep testing, (c) as-tempered Alloy D2, (d) Alloy D2 after creep testing, (e) as-tempered Alloy D3, (f) Alloy D3 after creep testing. TEM micrographs of (g) as-tempered Alloy D3 and (h) Alloy D3 after creep testing. (i) EDX result of Laves phase. (Distributions of Cr, Nb, Mo and W are attached in the right of microstructure of Alloys D1-D3 after creep testing).

4. Discussion

4.1. Extrapolation of the PM-TR-BCNN framework

The alloy design results highlight the limitations of prediction frameworks that lack uncertainty modeling, as they tend to yield excessively high and unreliable results. In contrast, the PM-TR-BCNN framework developed in this study produces more reliable and realistic predictions. To further validate the extrapolation capabilities of this framework, two types of external creep data from the Cambridge University Phase Transformation Group [55] were tested: (1) creep life data for heat-resistant steels tested under stress levels higher than the maximum value in this study's 650 °C subset (i.e., over 180 MPa) and (2) Creep life data of steels that are not part of this study's dataset but have elemental compositions within the upper and lower limits of the elements in the dataset. These external data enabled the evaluation of the predictive performance of the PM-TR-BCNN framework against that of the PM-TR-CNN framework (without uncertainty modeling) and a basic CNN framework (lacking both uncertainty and PM-TR modules).

Fig. 9(a) presents the prediction results from all three frameworks for high-stress tested heat-resistant steels. Under either the PM-TR-CNN or basic CNN framework (performance of basic CNN model shown in Table S4), the predicted creep lives for most samples are significantly below their experimental lives. In contrast, the PM-TR-BCNN framework provided highly accurate predictions with minimal standard deviations, as shown in Fig. 9(b). This demonstrates the superior extrapolation capability and stability of the uncertainty-based framework.

Fig. 10 further illustrates the predictions for heat-resistant steels with diverse compositions. The PM-TR-BCNN framework exhibits the highest accuracy and stability, validating its ability to reliably predict the creep lives of steels with previously untested compositions and conditions. This performance surpasses that of frameworks without uncertainty modeling, emphasizing the significance of the framework in optimizing heat-resistant steel composition and processing parameters. Overall, the uncertainty-based model enhances the reliability of alloy assessments, which are crucial for advancing material design and ensuring robust validation for practical applications.

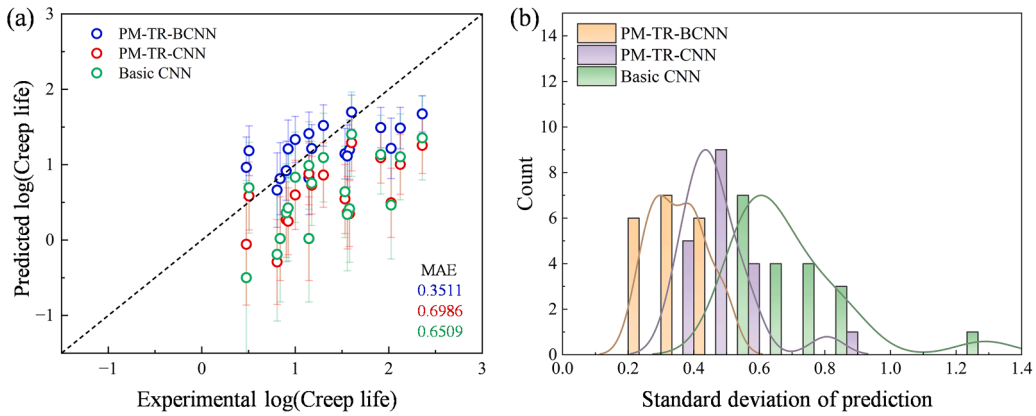

Fig. 9. Creep-life prediction results from different frameworks at stresses outside the dataset used for training the model. (a) Predicted versus experimental creep life. (b) Standard deviation distribution of the predicted creep lives.

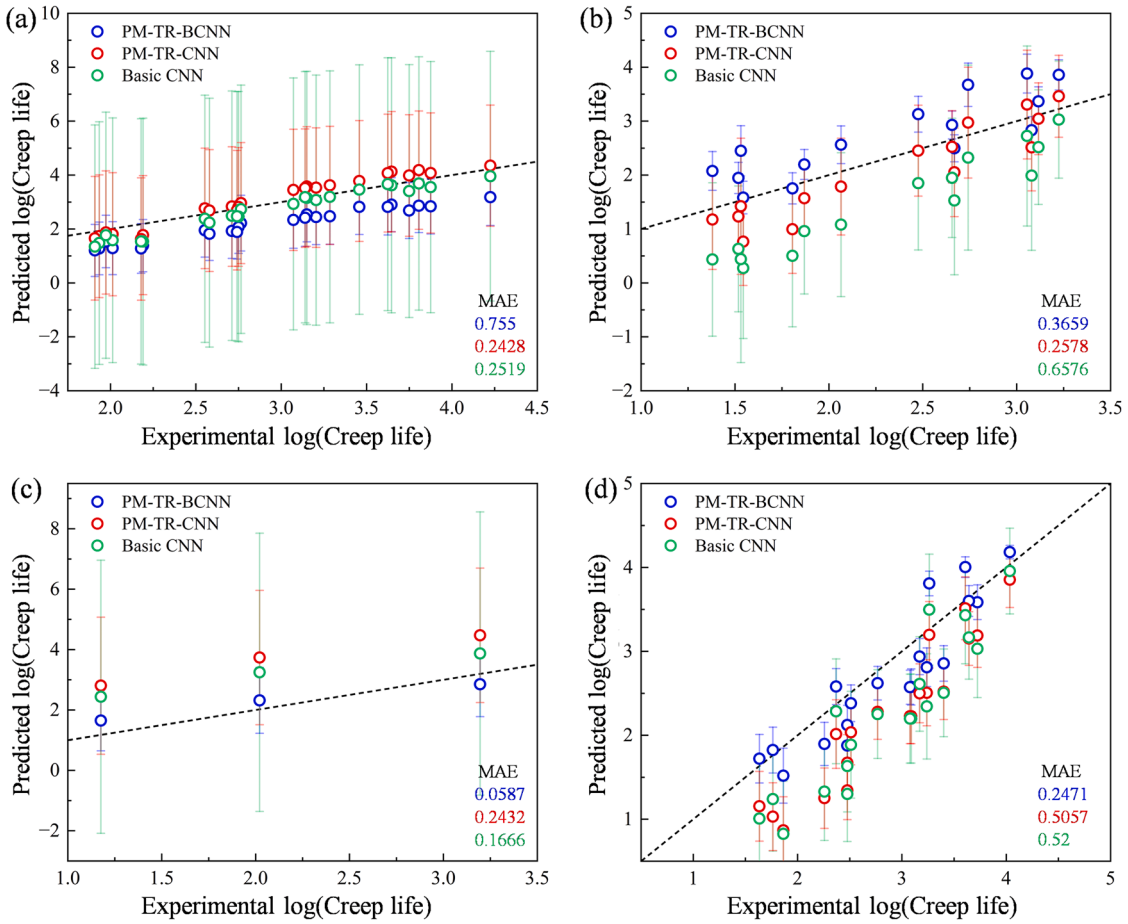

Fig. 10. Predicted creep life results from different frameworks versus experimental values for heat-resistant steels with different compositions: (a) 12Cr-1Mo-1 W, (b) 12Cr-Mo, (c) 11.5Cr-Mo, and (d) 10Cr-2Mo.

4.2. The effect of PM-TR module on creep life prediction

The PM-TR-BCNN model developed here accurately predicts the creep life and evaluates the prediction reliability, with uncertainty playing a key role in alloy design. This section further examines the contribution of the PM and TR modules to the prediction of creep life, particularly their impact on the performance of a model with small sample sizes.

Based on the subset with test temperature of 650 °C, Fig. 11 compares the prediction accuracy of the PM–TR–BCNN and basic BCNN models on a validation set using training datasets with different numbers of samples. The results demonstrate that the PM-TR-BCNN model consistently outperforms the basic BCNN model for all sizes of training dataset, with a particularly noticeable improvement when the dataset contains fewer samples. For example, when the training set contained only 24 samples, the basic BCNN achieved an accuracy of 69.1 ± 15.3 %, while incorporating the PM and TR modules boosted the accuracy to 75.3 ± 15.9 %. As the size of the training set increases, both models

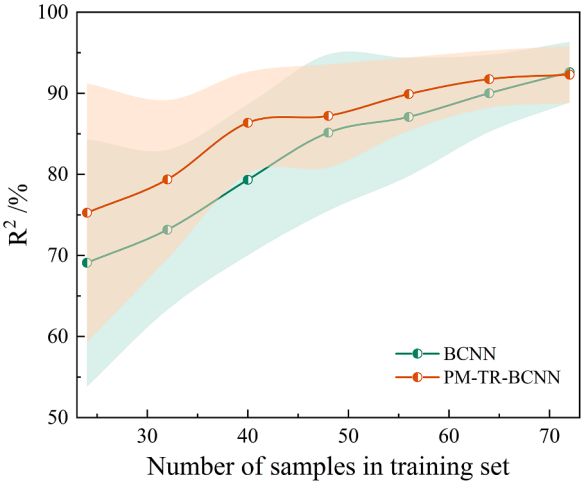

Fig. 11. Comparison of R^2 between PM-TR-BCNN and BCNN models given different numbers of datapoints used for training.

exhibit improvements in accuracy and a reduction in error bars. When the training set contained 72 samples, the accuracies of both models were similar. Additionally, the PM-TR-BCNN model exhibited smaller prediction error bars in most cases, which means that the model is more stable. The individual contributions of the PM and TR modules were quantified through ablation studies using three variants of the model (the training set contained 90 samples): the full PM-TR-BCNN, PM-BCNN (without the TR module), and TR-BCNN (without the PM module). Evaluation on the creep life validation set with R^2 metric revealed that the removal of the TR module reduced R^2 from 95.5 % to 95.3 %, while the exclusion of the PM module resulted in a more significant decrease to 94 %.

These findings underscore the important role of the PM and TR modules in improving the accuracy and stability of the model. The integration of these modules enhances the ability of the model to maintain reliable performance, even with limited data. This is particularly important in applications such as alloy design, where sample sizes are limited but precision and reliability are critical.

4.3. Incorporation of transformation-related features

In addition to PF, martensitic transformation is a fundamental aspect of creep-resistant steels, yet conventional feature sets often fail to explicitly capture it. Inspired by the work of Peng et al. [56], who demonstrated strong correlations between Ms, A3 temperatures, and tensile strength in heat-resistant alloys, we calculated the Ms and A3 temperatures for each alloy in our dataset using Thermo-Calc and incorporated them into the feature set. Their association with the tensile properties and creep life involved in this study was further evaluated. Both Ms and A3 demonstrated moderate PCC and MIC scores with respect to tensile properties (including YS, UTS, TEL, and RA) as well as creep life, confirming their relevance across multiple mechanical properties (Figs. S13-S15). Ms and A3 were further incorporated into the PM-TR-BCNN model to evaluate their impact on predictive performance. Although the overall accuracy of the model did not significantly improve (Fig. 12(a) and (b)), the inclusion of these features led to a certain reduction in predictive uncertainty (Fig. 12(c)). This indicates that transformation-related features offer valuable supplementary information, particularly in data-sparse regions where the model tends to extrapolate.

These findings suggest that Ms and A3, while not dominant predictors of performance, contribute to enhanced model robustness. Their inclusion reflects the importance of integrating physically meaningful features in data-driven materials design. Future work may explore more refined transformation descriptors or physics-informed neural network architectures to further leverage domain-specific metallurgical knowledge.

4.4. Function of uncertainty in alloy design

This section discusses the crucial role of uncertainty in alloy design. To illustrate this, the single-objective design results focusing on creep life and the dual-objective design results considering both creep life and uncertainty were compared based on all PM-TR-BCNN models (with accuracy greater than 95 %). For reference, PM-TR-CNN models without uncertainty modeling were also evaluated. The Mo–W content distributions of the alloys designed under these three scenarios are shown in Fig. 13(a). The results show that alloys designed using predictions, including uncertainty, primarily fall within a region characterized by a low Mo content (0.4–1 wt. %) and high W content (>1.7 wt. %), which aligns with the ideal Mo/W ratio range given by the composition limits of the dataset. In contrast, when uncertainty was not considered and the PM-TR-BCNN model was used with creep life as the sole objective, the resulting alloys were found in a region with high Mo and low W contents. This design is deemed unreliable because the Mo/W ratio is not

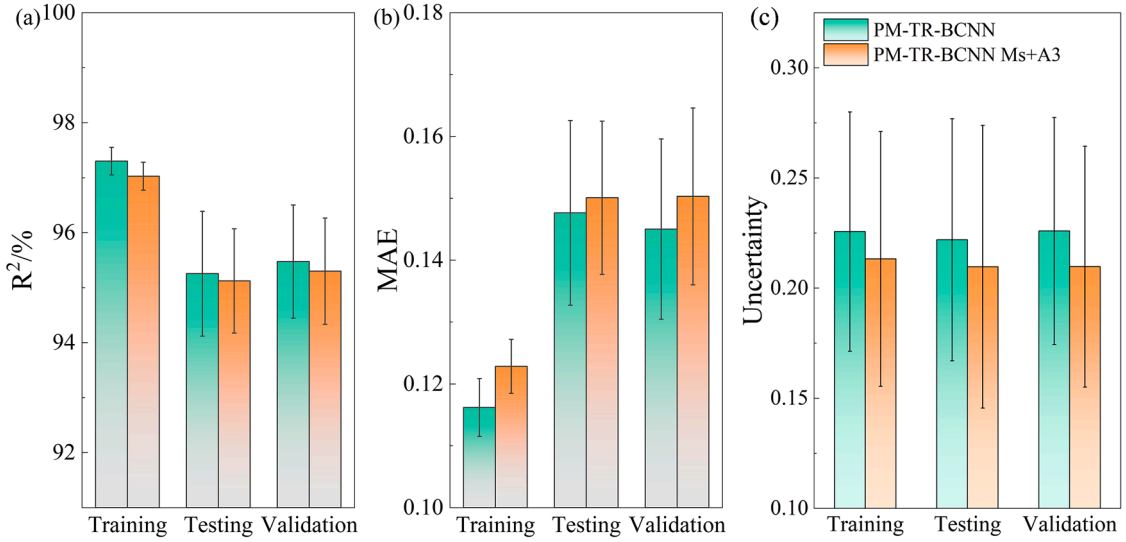

Fig. 12. Performance comparison of PM-TR-CNN model before and after introducing A3 and Ms. (a) R2. (b) MAE. (c) Prediction uncertainty.

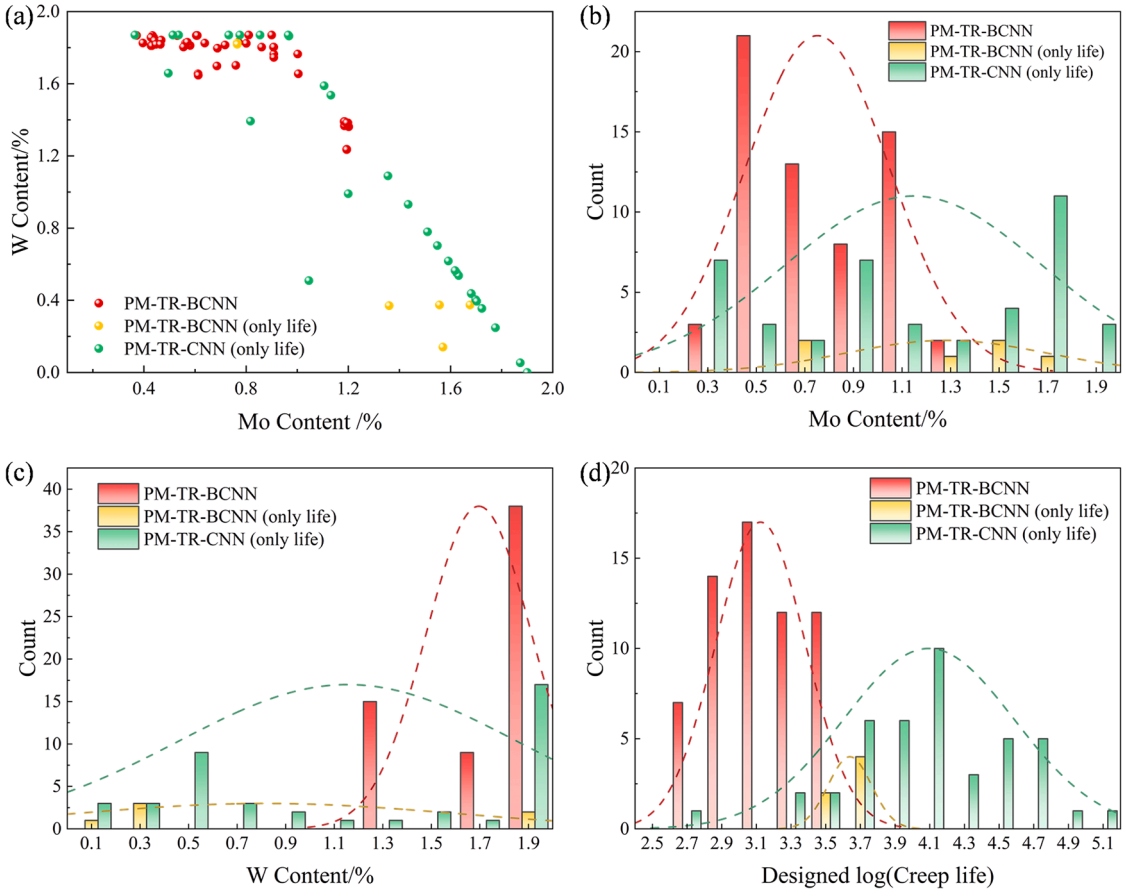

Fig. 13. Comparisons of the design results based on PM-TR-BCNN and PM-TR-CNN models. (a) Mo-W space. (b) Mo distribution. (c) W distribution. (d) Creep-life distribution.

optimal for creep resistance. Additionally, when the PM-TR-CNN model (without uncertainty modeling) was used for single-objective creep-life prediction, the designed alloys were predominantly distributed along the 1.9 % Mo equivalent constraint line. This resulted in a scattered Mo–W distribution, providing little useful insight into further alloy optimization. These findings highlight the superiority of the PM-TR-BCNN model with a dual-objective optimization for identifying the most promising alloy solutions with a greater efficiency.

Further analyses of the Mo and W content distributions of the designed alloys are presented in Fig. 13(b) and (c), respectively. The PM-TR-BCNN dual-objective results show narrower distributions of Mo and W contents compared to the results of the other models, highlighting the value of incorporating uncertainty in the design process. The predicted creep-life distributions of the alloys for each design scenario are shown in Fig. 13(d). The PM-TR-CNN model predicted unrealistically long creep lives, while the PM-TR-BCNN model predicted significantly shorter creep lives, which are considered more reliable and realistic. However, the creep-life values predicted by single-objective designs are high, with considerable uncertainty values, which introduce significant risks when validating alloys. When uncertainty was incorporated as part of the optimization objective, the predicted creep life closely matched the experimental results. The collective results demonstrate the essential role of inclusion of uncertainty in alloy design to identify compositions with high hypothetical creep resistance and ensure that the performance prediction is sufficiently reliable for practical applications.

In addition, it was found in the PM-TR-BCNN based design that higher predicted creep lives are typically associated with increased uncertainty. This trend can be considered from several perspectives. First, under specific test conditions (e.g., 650 °C and 140 MPa), the longer predicted creep lives often correspond to regions of the design space that are sparsely represented in the training dataset. This is because the genetic algorithm used in this study is designed to search for compositions/processing solutions with superior performance, inherently driving the model to extrapolate beyond its empirical experience. Such extrapolative predictions naturally carry higher epistemic uncertainty due to limited data support. Second, it was found that the PF factor tends to exhibit higher predictive uncertainty for alloys associated with longer creep lives. This variability reflects the increased difficulty in accurately modeling the strengthening contributions of these high-performance compositions. The underlying reason is likely the presence of complex precipitation behaviors that are not fully captured by the current set of features, especially in data-sparse regions of the design space. Third, beyond feature-level limitations, there is also greater intrinsic variability in the material behavior itself at longer timescales. Creep deformation under prolonged exposure becomes increasingly governed by time-dependent mechanisms such as grain boundary sliding, void nucleation and growth, and the coarsening or transformation of precipitates. These processes are highly sensitive to microstructural conditions and can vary significantly between nominally similar samples. As such, they introduce fundamental uncertainty that is challenging to capture, particularly when only indirect or limited descriptors are available. This materials-side complexity further contributes to the elevated uncertainty seen in predictions of long-term creep rupture. Overall, this increased uncertainty emphasizes the critical importance of incorporating uncertainty quantification in the alloy design process.

4.5. *Challenges and future work*

While the proposed PM-TR-BCNN framework demonstrates

promising capabilities by integrating physical knowledge with uncertainty-aware machine learning for alloy design, several simplifications and limitations remain.

First, the PF module currently ignores the influence of Laves phase kinetics, which may play a key role in long-term creep degradation, particularly under high-temperature or extended service conditions. Addressing this limitation through CALPHAD-based phase simulations or experimental calibration could enhance prediction. Similarly, although the PM and TR modules incorporate domain-informed variables, they are ultimately based on empirical simplifications and do not fully capture the complexity of microstructural evolution mechanisms, such as grain coarsening, elemental segregation, and secondary phase transformation. These aspects represent important directions for improving the physical interpretability and accuracy of the model.

Second, the differences between predicted and experimental creep lives may partly arise from limitations in the model itself, but could also reflect possible inconsistency in sample fabrication. To investigate this, a computational approach was used as a means of evaluation. For the designed alloys D1–D3, the five most similar alloys from the dataset were identified based on Euclidean distance in the normalized compositional space. The differences in both predicted and experimental creep lives were then compared between each designed alloy and its nearest neighbors. As shown in Fig. S16, a consistent trend emerged: greater compositional distance was associated with larger differences in both predicted and experimental creep life. This suggests that the observed discrepancies are not the result of inconsistencies in sample fabrication, but instead reflect the expected variability when making predictions beyond the range of training data. These findings also confirm that the designed alloys are compositionally and metallurgically aligned with the NIMS alloy system.

Nevertheless, fabrication-related factors can still significantly influence creep performance and cannot be ignored. Even small deviations in alloy composition, the presence of undetected secondary phases, or slight variations in testing conditions can affect the results. These subtle factors are difficult to fully capture in existing datasets. To explore the impact of such uncertainties, a local sensitivity analysis was conducted for alloy E1. Key input variables, including stress and selected alloying elements, were varied by ± 10 %, and the corresponding effects on predicted creep life were assessed (Fig. S17). The results show that the model is particularly sensitive to variations in test conditions and certain elements, such as carbon and manganese. This indicates that even modest experimental inaccuracies in these parameters could lead to significant prediction errors. Additionally, uncertainty propagation was performed via 1000 samples generated from normally distributed input features (Fig. S18). The results reveal that small fluctuations in input values can lead to significant variability in predicted creep life. Together, these findings highlight the difficulty in separating model bias from experimental noise and underscore the value of incorporating uncertainty-aware methods in materials design.

Data quality, therefore, imposes a fundamental upper bound on the predictive performance of machine learning models in materials science. Improving data fidelity is essential for enhancing both accuracy and uncertainty quantification. To address this, a set of strategies including dimensionality reduction, targeted sample augmentation, and the integration of domain knowledge into the modeling workflow have been outlined [57]. Building on these insights, a synergistic data quality governance framework can be proposed for future development of our modeling pipeline. These measures aim to enhance model robustness and reliability in practical alloy design applications. In the long term, the goal is to represent experimental tolerances through probabilistic input distributions and validate predictions against well-controlled laboratory datasets. These steps will allow us to decouple epistemic uncertainty (model-related) from aleatoric uncertainty (data-related), thereby enhancing the robustness, interpretability, and reliability of ML-guided alloy design.

5. Conclusion

To provide an effective method for creep life prediction of metallic materials and creep life-oriented alloy design, a Bayesian convolutional neural network framework (PM-TR-BCNN) integrating physical metallurgy and transfer learning is proposed, which is based on uncertainty modeling. The conclusions are as follows:

(1) The uncertainty-based ML framework significantly enhanced the predictive performance for creep life, particularly in terms of accuracy and reliability. The framework provides accurate creep life predictions with uncertainty quantification. In addition, the framework demonstrated strong extrapolation capabilities, delivering dependable results even under conditions outside the training dataset. Consequently, the inclusion of uncertainty enhances both the precision of the predictions and the robustness of the material property assessments, ensuring a realistic performance evaluation for practical applications.
(2) The PM-TR-BCNN framework was combined with a genetic algorithm and used for creep-life-directed alloy design. The PM-TR-BCNN framework showed better predictive design reliability, especially for the Mo and W contents, compared to alloy design without uncertainty. The improved creep life of the alloys designed by the PM-TR-BCNN model was validated experimentally, confirming the important role of uncertainty-guided alloy design.
(3) The PM and TR modules are essential for enhancing model performance, particularly in scenarios with limited data. The predictive accuracy of the proposed model was significantly improved by integrating these modules into a BCNN to include domain-specific knowledge, enabling better generalization and more reliable predictions compared to similar models.

CRediT authorship contribution statement

Chenchong Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xiaolu Wei:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Resources, Project administration, Methodology, Funding acquisition, Formal analysis. **Sybrand van der Zwaag:** Writing – review & editing, Formal analysis, Conceptualization. **Qiang Wang:** Conceptualization, Formal analysis, Writing – review & editing. **Wei Xu:** Writing – review & editing, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The research was financially supported by National Key R&D Program (No. 2024YFB3713705). The financial support provided by National Natural Science Foundation of China (Nos. 52311530082, U22A20106 and 52404395), Postdoctoral Fellowship Program of CPSF under Grant Number GZC20240222, the China Postdoctoral Science Foundation (2024M750370) and Xingliao Talent Program (Grant No. XLYC2203027) is gratefully acknowledged. The authors gratefully acknowledge the financial support by the Fundamental Research Funds for the Central Universities (N2425034). This research was conducted (in part) using the creep data sheet (CDS) provided by the Materials Data Platform (MDPF) of the National Institute for Materials Science (NIMS).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.actamat.2025.121073.

References

[1] S.H. Ryu, J. Yu, A new equation for the Cr equivalent in 9 to 12 pct Cr steels, Metall. Mater. Trans. A 29 (1998) 1573–1578, https://doi.org/10.1007/s11661-998-0080-7.

[2] T. Fujita, T. Sato, N. Takahashi, Effect of Mo and W on long term creep rupture strength of 12 %Cr heat-resisting steel containing V, Nb and B, Trans. Iron Steel Inst. Jpn. 18 (1978) 115–124, https://doi.org/10.2355/isijinternational1966.18.115.

[3] Q. Lu, W. Xu, S. van der Zwaag, A strain-based computational design of creep-resistant steels, Acta Mater 64 (2014) 133–143, https://doi.org/10.1016/j.actamat.2013.10.004.

[4] Q. Lu, W. Xu, S. van der Zwaag, The design of a compositionally robust martensitic creep-resistant steel with an optimized combination of precipitation hardening and solid-solution strengthening for high-temperature use, Acta Mater 77 (2014) 310–323, https://doi.org/10.1016/j.actamat.2014.06.007.

[5] H. Basoalto, S.K. Sondhi, B.F. Dyson, M. McLean, A generic microstructure-explicit model of creep in nickel-base superalloys, in: K.A. Green, T.M. Pollock, H. Harada, T.E. Howson, R.C. Reed, J.J. Schirra, S. Walston (Eds.), Superalloys 2004, Minerals, Metals & Materials Soc, Warrendale, 2004, pp. 897–906, https://doi.org/10.7449/2004/Superalloys_2004_897_906.

[6] T. Goswami, Development of generic creep–fatigue life prediction models, Mater. Des. 25 (2004) 277–288, https://doi.org/10.1016/j.matdes.2003.11.001.

[7] S. Holmström, P. Auerkari, A robust model for creep-fatigue life assessment, Mater. Sci. Eng. A 559 (2013) 333–335, https://doi.org/10.1016/j.msea.2012.08.107.

[8] J. Mackerle, Creep and creep fracture/damage finite element modelling of engineering materials and structures: an addendum, Int. J. Pressure Vessels Piping 81 (2004) 381–392, https://doi.org/10.1016/j.ijpvp.2004.03.007.

[9] Q. Meng, Z. Wang, Creep damage models and their applications for crack growth analysis in pipes: a review, Eng. Fract. Mech. 205 (2019) 547–576, https://doi.org/10.1016/j.engfracmech.2015.09.055.

[10] V.S. Srinivasan, J. Vanaja, B.K. Choudhary, K. Laha, Modeling of creep deformation behaviour of RAFM steel, Trans. Indian Inst. Met. 69 (2016) 567–571, https://doi.org/10.1007/s12666-015-0812-5.

[11] M. Tehami, K.E. Ramdane, Creep behaviour modelling of a composite steel–concrete section, J. Constr. Steel Res. 65 (2009) 1029–1033, https://doi.org/10.1016/j.jcsr.2009.01.001.

[12] N. Torić, A. Harapin, I. Boko, Modelling of steel creep at high temperatures using an implicit creep model, Key Eng. Mater. 553 (2013) 13–22, https://doi.org/10.4028/www.scientific.net/KEM.553.13.

[13] W. Yan, W. Wang, Y. Shan, K. Yang, W. Sha, 9-12Cr Heat-Resistant Steels, Springer, 2015.

[14] F.R. Larson, J. Miller, A time-temperature relationship for rupture and creep stresses, Trans. ASME 74 (1952) 765–771, https://doi.org/10.1115/1.4015909.

[15] S.G.R. Brown, R.W. Evans, B. Wilshire, A comparison of extrapolation techniques for long-term creep strain and creep life prediction based on equations designed to represent creep curve shape, Int. J. Pressure Vessels Piping 24 (1986) 251–268, https://doi.org/10.1016/0308-0161(86)90125-0.

[16] R.W. Evans, J.D. Parker, B. Wilshire, The θ projection concept—a model-based approach to design and life extension of engineering plant, Int. J. Pressure Vessels Piping 50 (1992) 147–160, https://doi.org/10.1016/0308-0161(92)90035-E.

[17] R.A. Arutyunyan, S.V. Petrov, A.G. Tereshchenko, The θ-concept in the theory of creep and its modification for describing the creep limit, Strength Mater 21 (1989) 1613–1617, https://doi.org/10.1007/BF01533399.

[18] A.A. Ayubali, A. Singh, B.P. Shanmugavel, K.A. Padmanabhan, A phenomenological model for predicting long-term high temperature creep life of materials from short-term high temperature creep test data, Int. J. Mech. Sci. 202-203 (2021) 106505, https://doi.org/10.1016/j.ijmecsci.2021.106505.

[19] B. Wilshire, P.J. Scharning, R. Hurst, A new approach to creep data assessment, Mater. Sci. Eng. A 510-511 (2009) 3–6, https://doi.org/10.1016/j.msea.2008.04.125.

[20] B. Wilshire, P.J. Scharning, Long-term creep life prediction for a high chromium steel, Scripta Mater 56 (2007) 701–704, https://doi.org/10.1016/j.scriptamat.2006.12.033.

[21] B. Wilshire, P.J. Scharning, A new methodology for analysis of creep and creep fracture data for 9–12 % chromium steels, Int. Mater. Rev. 53 (2008) 91–104, https://doi.org/10.1179/174328008X254349.

[22] Z. Abdallah, K. Perkins, S. Williams, Advances in the Wilshire extrapolation technique—Full creep curve representation for the aerospace alloy Titanium 834, Mater. Sci. Eng. A 550 (2012) 176–182, https://doi.org/10.1016/j.msea.2012.04.054.

[23] M.T. Whittaker, W.J. Harrison, R.J. Lancaster, S. Williams, An analysis of modern creep lifing methodologies in the titanium alloy Ti6-4, Mater. Sci. Eng. A 577 (2013) 114–119, https://doi.org/10.1016/j.msea.2013.03.030.

[24] W. Ye, X. Hu, Y. Song, The relationship between creep and tensile properties of a nickel-based superalloy, Mater. Sci. Eng. A 774 (2020) 138847, https://doi.org/10.1016/j.msea.2019.138847.

[25] L.M. Kachanov, On rupture time under condition of creep, Int. J. Fract. 97 (1999) 11–18, https://doi.org/10.1023/A:1018671022008.

[26] Y.N. Rabotnov, Creep problems in structure members, J. Appl. Mech. 37 (1970) 249, https://doi.org/10.1115/1.3408479.

[27] Y. Liu, S. Murakami, Damage localization of conventional creep damage models and proposition of a new model for creep damage analysis, JSME Int. J. 41 (1998) 57–65, https://doi.org/10.1299/jsmea.41.57.

[28] S. Murakami, Y. Liu, M. Mizuno, Computational methods for creep fracture analysis by damage mechanics, Comput. Methods Appl. Mech. Eng. 183 (2000) 15–33, https://doi.org/10.1016/s0045-7825(99)00209-1.

[29] S. Murakami, N. Ohno, A continuum theory of creep and creep damage, in: A. Ponter, D. Hayhurst (Eds.), Creep in Structures, Springer-Verlag, Berlin, 1981, pp. 422–444, https://doi.org/10.1007/978-3-642-81598-0_28.

[30] S. Murakami, Notion of continuum damage mechanics and its application to anisotropic creep damage theory, J. Eng. Mater. Technol. ASME 105 (1983) 99–105, https://doi.org/10.1115/1.3225633.

[31] J.L. Chaboche, Anisotropic creep damage in the framework of continuum damage mechanics, Nucl. Eng. Des. 79 (1984) 309–319, https://doi.org/10.1016/0029-5493(84)90046-3.

[32] J.-L. Chaboche, Continuous damage mechanics — A tool to describe phenomena before crack initiation, Nucl. Eng. Des. 64 (1981) 233–247, https://doi.org/10.1016/0029-5493(81)90007-8.

[33] X. Wei, S. van der Zwaag, Z. Jia, C. Wang, W. Xu, On the use of transfer modeling to design new steels with excellent rotating bending fatigue resistance even in the case of very small calibration datasets, Acta Mater 235 (2022) 118103, https://doi.org/10.1016/j.actamat.2022.118103.

[34] C. Shen, C. Wang, X. Wei, Y. Li, S. van der Zwaag, W. Xu, Physical metallurgy-guided machine learning and artificial intelligent design of ultrahigh-strength stainless steel, Acta Mater 179 (2019) 201–214, https://doi.org/10.1016/j.actamat.2019.08.033.

[35] C. Wang, K. Zhu, P. Hedström, Y. Li, W. Xu, A generic and extensible model for the martensite start temperature incorporating thermodynamic data mining and deep learning framework, J. Mater. Sci. Technol. 128 (2022) 31–43, https://doi.org/10.1016/j.jmst.2022.04.014.

[36] Y. Liu, J. Wu, Z. Wang, X. Lu, M. Avdeev, S. Shi, C. Wang, T. Yu, Predicting creep rupture life of Ni-based single crystal superalloys using divide-and-conquer approach based machine learning, Acta Mater 195 (2020) 454–467, https://doi.org/10.1016/j.actamat.2020.05.001.

[37] J. Wang, Y. Fa, Y. Tian, X. Yu, A machine-learning approach to predict creep properties of Cr–Mo steel with time-temperature parameters, J. Mater. Res. Technol. 13 (2021) 635–650, https://doi.org/10.1016/j.jmrt.2021.04.079.

[38] D. Shin, Y. Yamamoto, M.P. Brady, S. Lee, J.A. Haynes, Modern data analytics approach to predict creep of high-temperature alloys, Acta Mater 168 (2019) 321–330, https://doi.org/10.1016/j.actamat.2019.02.017.

[39] C. Wang, X. Wei, D. Ren, X. Wang, W. Xu, High-throughput map design of creep life in low-alloy steels by integrating machine learning with a genetic algorithm, Mater. Des. 213 (2022) 110326, https://doi.org/10.1016/j.matdes.2021.110326.

[40] A. Zare, R.K. Hosseini, A breakthrough in creep lifetime prediction: leveraging machine learning and service data, Scripta Mater 245 (2024) 116037, https://doi.org/10.1016/j.scriptamat.2024.116037.

[41] D. Ren, C. Wang, X. Wei, Q. Lai, W. Xu, Building a quantitative composition-microstructure-property relationship of dual-phase steels via multimodal data mining, Acta Mater 252 (2023) 118954, https://doi.org/10.1016/j.actamat.2023.118954.

[42] D. Ren, C. Wang, X. Wei, Y. Zhang, S. Han, W. Xu, Harmonizing physical and deep learning modeling: a computationally efficient and interpretable approach for property prediction, Scripta Mater 255 (2025) 116350, https://doi.org/10.1016/j.scriptamat.2024.116350.

[43] Y. Li, H. Li, C. Wang, P.E. Jose Rivera-Diaz-del-Castillo, The role of physical metallurgical relationships in enhancing alloy properties prediction and design: a case study on Q&P steel, MGE Advances 3 (2025) e70, https://doi.org/10.1002/mgea.70.

[44] X. Geng, F. Wang, H.-H. Wu, S. Wang, G. Wu, J. Gao, H. Zhao, C. Zhang, X. Mao, Data-driven and artificial intelligence accelerated steel material research and intelligent manufacturing technology, MGE Advances 1 (2023) e10, https://doi.org/10.1002/mgea.10.

[45] S. Han, C. Wang, Y. Zhang, W. Xu, H. Di, Employing deep learning in non-parametric inverse visualization of elastic–plastic mechanisms in dual-phase steels, MGE Advances 2 (2024) e29, https://doi.org/10.1002/mgea.29.

[46] S. Xiang, X. Chen, Z. Fan, T. Chen, X. Lian, A deep learning-aided prediction approach for creep rupture time of Fe–Cr–Ni heat-resistant alloys by integrating textual and visual features, J. Mater. Res. Technol. 18 (2022) 268–281, https://doi.org/10.1016/j.jmrt.2022.02.099.

[47] X.-C. Zhang, J.-G. Gong, F.-Z. Xuan, A deep learning based life prediction method for components under creep, fatigue and creep-fatigue conditions, Int. J. Fatigue 148 (2021) 106236, https://doi.org/10.1016/j.ijfatigue.2021.106236.

[48] 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 (2002) 22–29, https://doi.org/10.1016/S0921-5093(01)01965-7.

[49] J. Peng, Y. Yamamoto, M.P. Brady, S. Lee, J.A. Haynes, D. Shin, Uncertainty quantification of machine learning predicted creep property of alumina-forming austenitic alloys, JOM 73 (2021) 164–173, https://doi.org/10.1007/s11837-020-04423-x.

[50] S.Y. Kwon, Y. Yamamoto, J. Peng, M.P. Brady, T.R. Watkins, J.A. Haynes, D. Shin, Physics-coupled data-driven design of high-temperature alloys, Acta Mater 284 (2025) 120622, https://doi.org/10.1016/j.actamat.2024.120622.

[51] NIMS, Creep Data Sheet (CDS). https://cds.nims.go.jp/, [1 November 2023].

[52] K. Sawada, K. Kimura, F. Abe, Y. Taniuchi, K. Sekido, T. Nojima, T. Ohba, H. Kushima, H. Miyazaki, H. Hongo, T. Watanabe, Catalog of NIMS creep data sheets, Sci. Technol. Adv. Mater. 20 (2019) 1131–1149, https://doi.org/10.1080/14686996.2019.1697616.
[53] A. Kelly, The strength of aluminium silver alloys, Philos. Mag. 3 (1958) 1472–1474, https://doi.org/10.1080/14786435808233338.
[54] Y. Tsuchida, K. Okamoto, Y. Tokunaga, Improvement of creep rupture strength of high Cr ferritic steel by addition of W, ISIJ Int 35 (1995) 317–323, https://doi.org/10.2355/isijinternational.35.317.
[55] Phase Transformation Group at University of Cambridge, MAP Program & Data Library Contents. https://www.phase-trans.msm.cam.ac.uk/map/map.html, [1 November 2023].
[56] J. Peng, Y. Yamamoto, J.A. Hawk, E. Lara-Curzio, D. Shin, Coupling physics in machine learning to predict properties of high-temperatures alloys, npj Comput. Mater. 6 (2020) 141, https://doi.org/10.1038/s41524-020-00407-2.
[57] Y. Liu, Z. Yang, X. Zou, S. Ma, D. Liu, M. Avdeev, S. Shi, Data quantity governance for machine learning in materials science, Nat. Sci. Rev. 10 (2023) nwad125, https://doi.org/10.1093/nsr/nwad125.