

<https://doi.org/10.1038/s41524-025-01906-w>

Alloy design integrating natural language processing and machine learning: breakthrough development of low-cost, high-performance Ni-based single-crystal superalloys



Jian Yao^{1,4}, Zi Wang^{1,4}, Juncheng Wang¹, Wanchan Yu², Yuxuan Chen², Weifu Li², Jianhui Wei³, Yunxing Zhao³, Yan Wang¹, Li Wang¹, Liming Tan¹✉, Lan Huang¹✉, Feng Liu¹✉ & Yong Liu¹

The traditional design of single-crystal superalloys relies heavily on trial-and-error experimentation, which is time-consuming and costly. Here, we present an intelligent alloy design strategy that integrates natural language processing (NLP) and machine learning (ML). A domain-specific NLP model was developed to automatically extract γ' solvus temperature data from scientific literature, enabling the construction of a high-quality database. Machine learning models trained on this data accurately predict both γ' solvus temperature and creep life. Guided by these models, we screened over 340000 virtual compositions and successfully designed a new low-cost alloy, CSU-S1.

Experimental validation shows that CSU-S1 achieves a γ' solvus temperature near 1300 °C and a creep life of 224.7 h at 1100 °C/137 MPa, comparable to third-generation single-crystal superalloys, while using only 3.1 wt% Re and costing just 121 USD/kg. This work not only delivers a high-performance, cost-effective superalloy but also demonstrates a generalizable “knowledge-to-innovation” design paradigm, offering a powerful new route to accelerate the development of advanced engineering materials.

Single crystal superalloys are indispensable in modern aerospace and power generation systems, where they serve as critical components in the hottest sections of jet engines and gas turbines^{1–3}. Their exceptional resistance to deformation at elevated temperatures arises from a microstructure dominated by coherent γ' precipitates embedded in a γ matrix. Nevertheless, the design of these alloys confronts formidable challenges, primarily encompassing compositional optimization and performance prediction^{4,5}. Conventional alloy design relies heavily on extensive experimental validation and empirical correlations, incurring considerable time and cost, particularly for critical properties such as solvus temperature, creep resistance, and oxidation resistance. Moreover, traditional methods become inefficient when optimizing complex compositions against multifaceted property targets. Hence, a new methodology capable of processing high-dimensional

data and delivering efficient prediction and optimization is imperative to accelerate alloy development⁶.

Among the many factors that influence high-temperature performance, the γ' solvus temperature has long been used as a practical indicator of alloy capability, effectively reflecting the alloy's heat tolerance. Thermo-dynamic modeling tools such as CALPHAD offer valuable estimates of equilibrium properties, including the γ' solvus temperature. However, their predictions depend on the completeness and accuracy of underlying databases and often require experimental calibration. At the same time, reliable experimental measurements of this property are available only for a limited set of legacy alloys and are scattered across decades of scientific literature. Manually collecting such data is time-consuming and does not scale to the needs of modern high-throughput design strategies.

¹State Key Laboratory of Powder Metallurgy, Central South University, Changsha, China. ²College of Informatics, Huazhong Agricultural University, Wuhan, China.

³Wan Ze Aviation Technology Co., LTD, Shenzhen, China. ⁴These authors contributed equally: Jian Yao, Zi Wang. ✉e-mail: limingtan@csu.edu.cn;

lhuan@csu.edu.cn; liufeng@csu.edu.cn

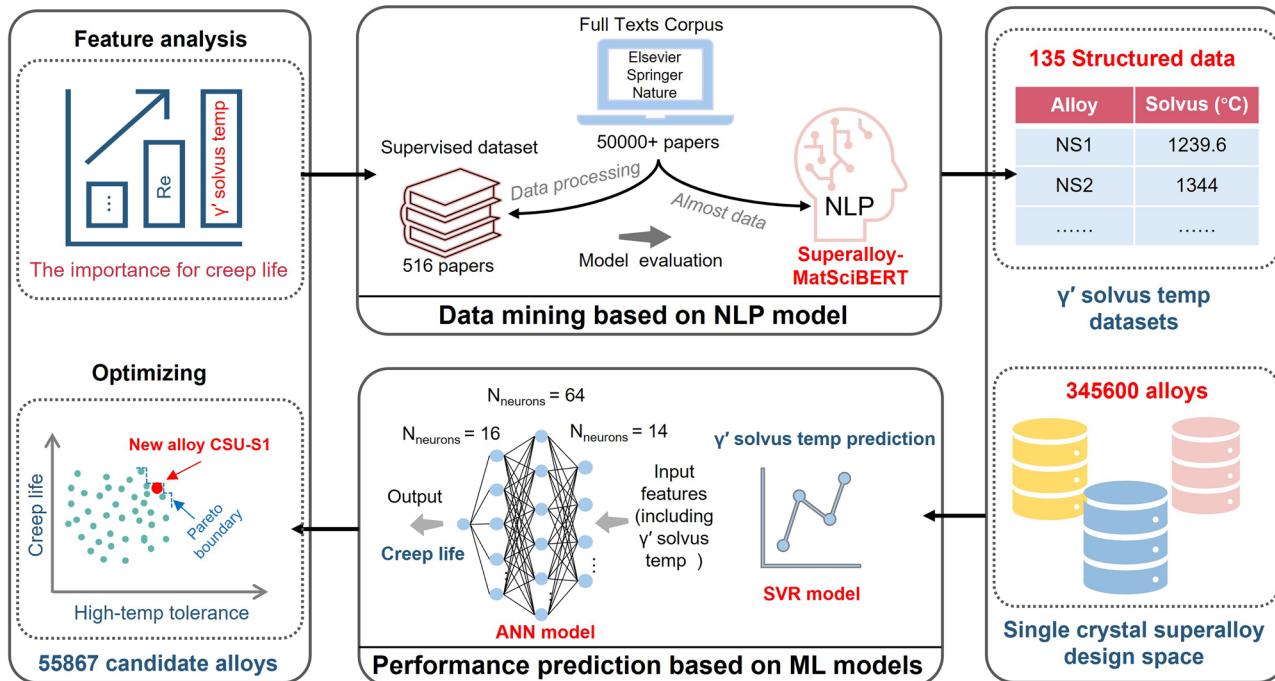


Fig. 1 | NLP and ML based design and prediction workflow for single-crystal superalloys.

The use of data-driven approaches to accelerate materials discovery has deep historical roots^{7–9}. Machine learning (ML) models, particularly artificial neural networks, learn complex nonlinearities from large alloy datasets, demonstrating strong potential in performance prediction and compositional optimization^{10–13}. Over thirty years ago, Bhadeshia et al.¹⁴ demonstrated that artificial neural networks could capture complex relationships between composition, processing, and mechanical behavior in steels, laying the conceptual foundation for modern machine learning applications in materials science. This vision has since evolved into successful computational alloy design frameworks that integrate thermodynamic modeling, machine learning, and multi-objective optimization. Notably, Conduit et al.¹⁵ and Menou et al.¹⁶ experimentally validated novel superalloys discovered through such integrated approaches, demonstrating the viability of computation-guided design.

Nevertheless, these strategies typically operate within restricted composition spaces and rely on curated datasets that are labor-intensive to assemble. For properties such as γ' solvus temperature, though central to design, they are reported sporadically across decades of literature in unstructured formats, making comprehensive data collection a major bottleneck. The emergence of natural language processing offers a transformative solution to this challenge¹⁷. By automatically extracting material properties from vast corpora of publications and patents, NLP enables the construction of structured, machine-readable databases that were previously unattainable. When combined with machine learning models capable of learning complex compositional dependencies, this NLP-driven data pipeline opens new pathways for high-throughput alloy design^{18,19}.

To overcome the data bottleneck that limits the scalability of computation-guided superalloy design, we present an end-to-end framework that transforms unstructured scientific text into actionable design knowledge. Using a domain-specific NLP model, we automatically extract experimentally measured γ' solvus temperatures from tens of thousands of publications and patents, constructing a large-scale, high-fidelity dataset that far exceeds the scope of manual curation. This dataset enables the training of machine learning models that accurately predict γ' solvus temperature and auxiliary predict creep life. Integrated into a multi-constraint high-throughput screening workflow, our approach evaluates over 340,000 virtual Ni-based single-crystal compositions, identifying promising candidates that balance heat tolerance, cost, density, microstructural stability, and

processability. By closing the loop between decades of fragmented literature and modern data-driven design, this work establishes a generalizable pathway for accelerating the discovery of next-generation high-performance alloys, as shown in Fig. 1.

Results

Rationale for a data-driven approach

Figure 2 summarizes the evolutionary trends of γ' solvus temperature and creep performance across six generations of single-crystal superalloys. With successive alloy generations, the γ' solvus temperature generally increases (Fig. 2a). However, there is no universal monotonic positive correlation between γ' solvus temperature and creep life. Several representative counterexamples highlight the limitations of relying solely on γ' solvus temperature to assess creep performance. For instance, PWA1480 exhibits a γ' solvus temperature approximately 35 °C higher than that of Rene N4, yet under testing conditions of 1100 °C and 137 MPa, it shows inferior creep life, primarily due to the absence of W element in PWA1480, which results in insufficient solid solution strengthening in the γ matrix²⁰. In contrast, TMS-238 and TMS-138 have nearly identical γ' solvus temperatures, but the former exhibits a creep life approximately twice that of the latter, owing to the higher partitioning ratios of refractory elements such as Re and Ru into the γ phase, thereby significantly enhancing the matrix's creep resistance. These observations indicate that the generational improvement in superalloy performance stems primarily from optimized elemental partitioning, particularly the enrichment of potent solid solution strengthening elements in the γ matrix, rather than merely increasing the volume fraction of the γ' phase.

Nevertheless, among the ten representative alloys shown in Fig. 2b, Fig. 2d illustrates a strong positive correlation ($R^2 = 0.94$) between γ' solvus temperature and creep life under the specific test condition of 1100 °C/137 MPa. It should be noted that although this statistical relationship is subject to certain biases, it provides useful design insight: γ' solvus temperature is better interpreted as a proxy for an alloy's high-temperature capability, i.e., its ability to maintain an effective volume fraction of γ' strengthening precipitates at elevated temperatures, rather than as a direct determinant of creep life. However, this does not imply that a higher γ' solvus temperature necessarily leads to better creep resistance. Previous studies²¹ have shown that although a higher solvus temperature enables the

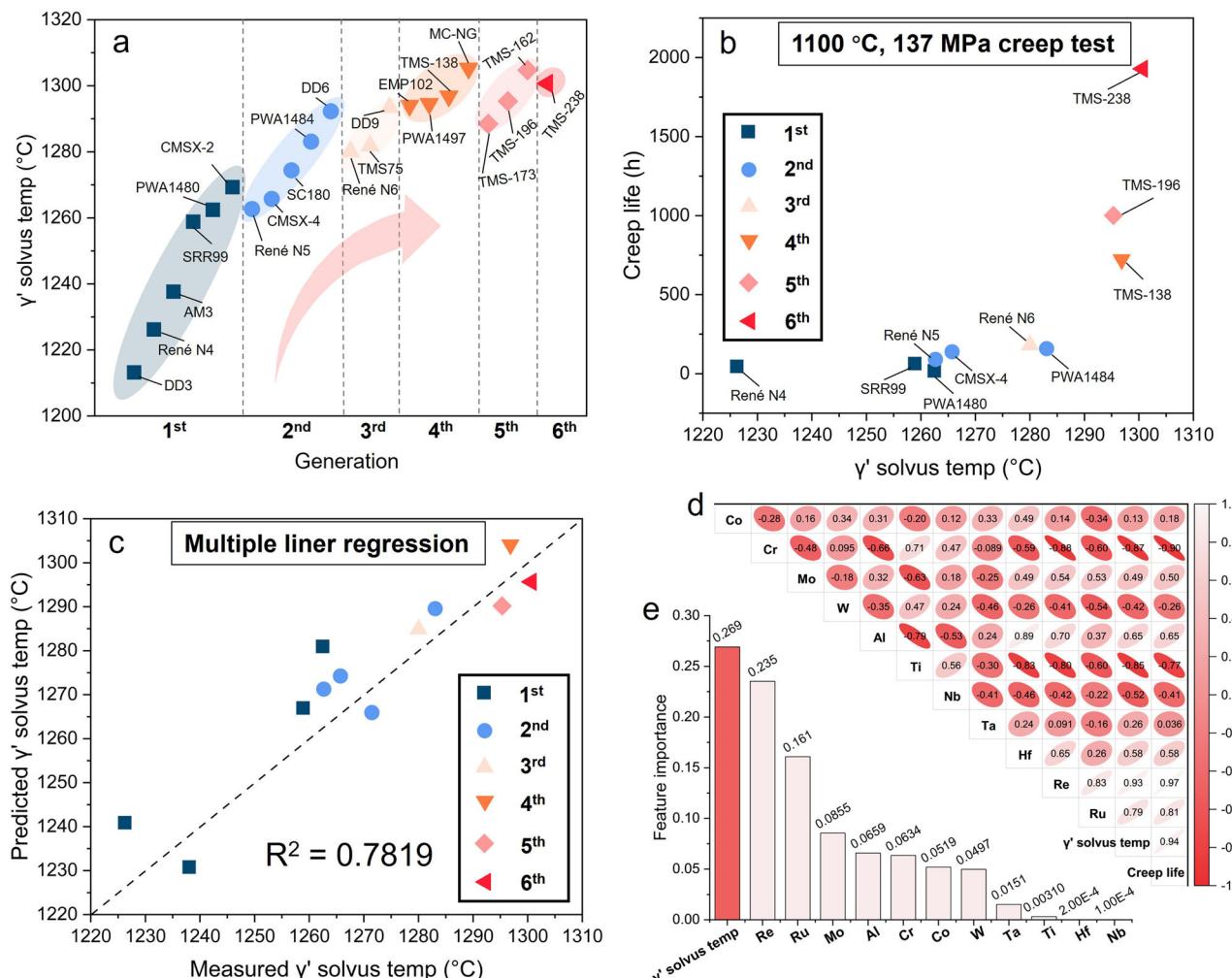


Fig. 2 | The effect of γ' solvus temperature on the creep life of single crystal superalloys of different generations. **a** Changes in the γ' solvus temperature of different generations of single-crystal superalloys. **b** The relationship between γ' solvus temperature and creep life under 1100°C, 137 MPa creep testing.

c Comparison of predicted γ' solvus temperature based on a multiple linear regression model with actual measured values, with an R^2 of 0.7819. **d** Correlation analysis of alloy elements, γ' solvus temperature, and creep life at 1100°C, 137 MPa. **e** Feature importance analysis was calculated using a random forest model.

retention of a greater volume fraction of γ' precipitates at elevated temperatures, exceeding a γ' volume fraction of ~70% may promote the formation of topologically close-packed (TCP) phases in the matrix. Therefore, this physical parameter must be carefully considered in alloy design. This understanding facilitates high-throughput alloy screening. Given that γ' solvus temperature is highly sensitive to composition and difficult to predict accurately using simple empirical models (the linear regression in Fig. 2c yields an R^2 of only 0.7819), its efficient acquisition represents a critical bottleneck in alloy design. To overcome this limitation, we employ a domain-specific natural language processing (NLP) model to automatically extract experimentally measured γ' solvus temperatures from tens of thousands of scientific publications and patents. This enables the creation of a large-scale, high-quality dataset that faithfully reflects how alloy composition influences the γ' solvus temperature, thereby providing a robust foundation for training accurate machine learning models and enabling high-throughput alloy design.

Superalloy-MatSciBERT framework

To address the scarcity of structured high-temperature alloy data, we developed Superalloy-MatSciBERT, a domain-adapted natural language processing framework that automatically extracts γ' solvus temperatures and other key properties from unstructured scientific literature. The framework operates as a three-stage pipeline²² (Fig. 3a-c): (i) sentence

classification (SC) to identify solvus-relevant passages; (ii) named entity recognition (NER) to extract alloy names, property types, and numerical values; and (iii) relation extraction (RE) to link entities into structured “alloy–property” pairs.

We harvested 52386 journal articles (2010–2023) from Elsevier, Springer, and Nature by querying the CrossRef REST API with the regex keyword “single [-] crystal.* superalloy” and DOI prefixes 10.1016/10.1007/10.1038, as seen in Table S3. An additional 532 single-crystal superalloy patents were downloaded as PDFs from Patent-Star via Selenium-controlled Chrome Driver in headless mode. All PDFs were converted to plain text; HTML tags, equations, and reference lists were removed with regular expressions. Sentences were segmented using NLTK Punkt augmented with a custom abbreviation dictionary, and temperature units were normalized to °C. The final corpus consists of 1.37 million sentences with an average length of 22.4 tokens. From 516 papers, two researchers performed independent and cross-validated annotation to produce three supervised datasets: a sentence-classification set of 1514 sentences (223 γ' -solvus relevant), a named-entity-recognition set of 256 sentences with character-level BIO tags for alloy names, property names, and temperature values, and a relation-extraction set of 152 sentences with [E1]/[E2] markers for alloy–temperature pairs. All data were randomly stratified at 6:2:2, with no single paper spanning splits, ensuring zero information leakage.

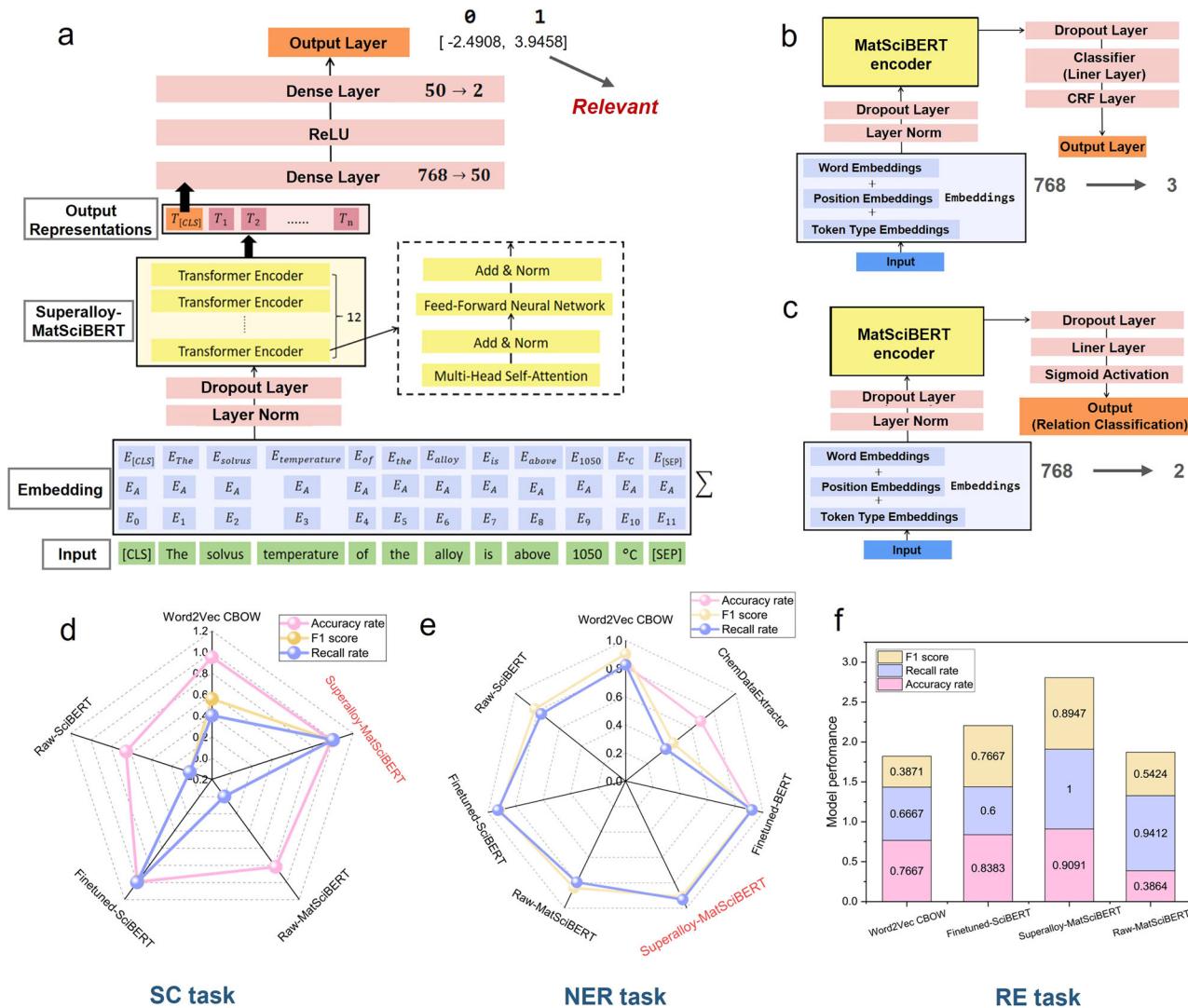


Fig. 3 | NLP-based alloy data extraction framework. **a**, a sentence classification (SC) model architecture. Input sentences are tokenized and tagged with special tokens [CLS] and [SEP]. A 12-layer Transformer encoder extracts deep semantic features, outputting binary classification probabilities. **b**, Named entity recognition model architecture. A conditional random field (CRF) layer precisely identifies entity boundaries, extracting key entities such as alloy names and solvus temperatures.

c, Relation extraction (RE) model architecture. Using entity markers [E1] and [E2], a fully connected layer with Sigmoid activation performs binary relation classification to construct “material–property” pairs. **d–f**, Model performance comparison. On the 516-paper annotated dataset, Superalloy-MatSciBERT outperforms other models in SC, NER, and RE, achieving superior F1 scores, accuracy, and recall.

Built on a 12-layer, 768-d MatSciBERT backbone, the framework was fine-tuned for three tasks: sentence classification feeds the [CLS] vector through Dense(50, ReLU) → Dense(2, Softmax) and uses Focal Loss ($\gamma = 2$) to counter the 1:6 imbalance; named-entity recognition appends a BiLSTM(128) + CRF layer after MatSciBERT to capture label dependencies; relation extraction inputs entity-marked sentences into MatSciBERT → Dropout(0.2) → Dense(2, Sigmoid).

To systematically evaluate the NLP framework, we built a manually annotated corpus of 516 papers selected from more than 50000 crawled documents, spanning diverse alloy systems and property parameters. All models were trained with AdamW (early stopping patience = 5, max length = 256), converging after 34 (SC), 100 (NER), and 103 (RE) epochs. On the test set, Superalloy-MatSciBERT achieved F1 scores of 0.990 (SC), 0.913 (NER), and 0.895 (RE), significantly outperforming raw SciBERT²³, MatSciBERT²⁴, and Word2Vec²⁵ baselines (Fig. 3d–f). Notably, in RE it attained F1 = 0.895, recall = 1, and accuracy = 0.909, confirming its superior ability to generate high-fidelity, machine-readable material data from text. This capability enabled the extraction of 135 experimental γ' solvus values—a critical dataset that underpins our high-throughput alloy design workflow.

Database Construction and Machine Learning Prediction Models
 Based on high-quality data automatically extracted from a vast corpus of literature using Superalloy-MatSciBERT, we have constructed a structured materials database tailored for single-crystal superalloys. This database includes 135 experimentally measured γ' solvus temperatures (Table S1) and 124 solidus temperatures (Table S2), along with their corresponding alloy compositions. It encompasses a wide range of complex alloys, including Ni-based and Co-based systems, providing a robust foundation for subsequent machine learning modeling. As shown in Fig. 4a, the γ' solvus temperatures are primarily concentrated within the 800–1350 °C range, while the solidus temperatures exhibit a broader distribution, reflecting significant differences in thermodynamic properties across various alloy systems. The compositional distributions (Fig. 4b, c) reveal that both datasets predominantly consist of Ni- and Co-based alloys, commonly containing key alloying elements such as Al, W, and Ta.

Using this database, we trained five regression models to predict the solidus temperature and γ' solvus temperature of alloys. Performance comparisons (Fig. 4d, e) indicate that the SVR.rbf model performs the best: on the test set, it achieves an R^2 of 0.949 for predicting the solidus

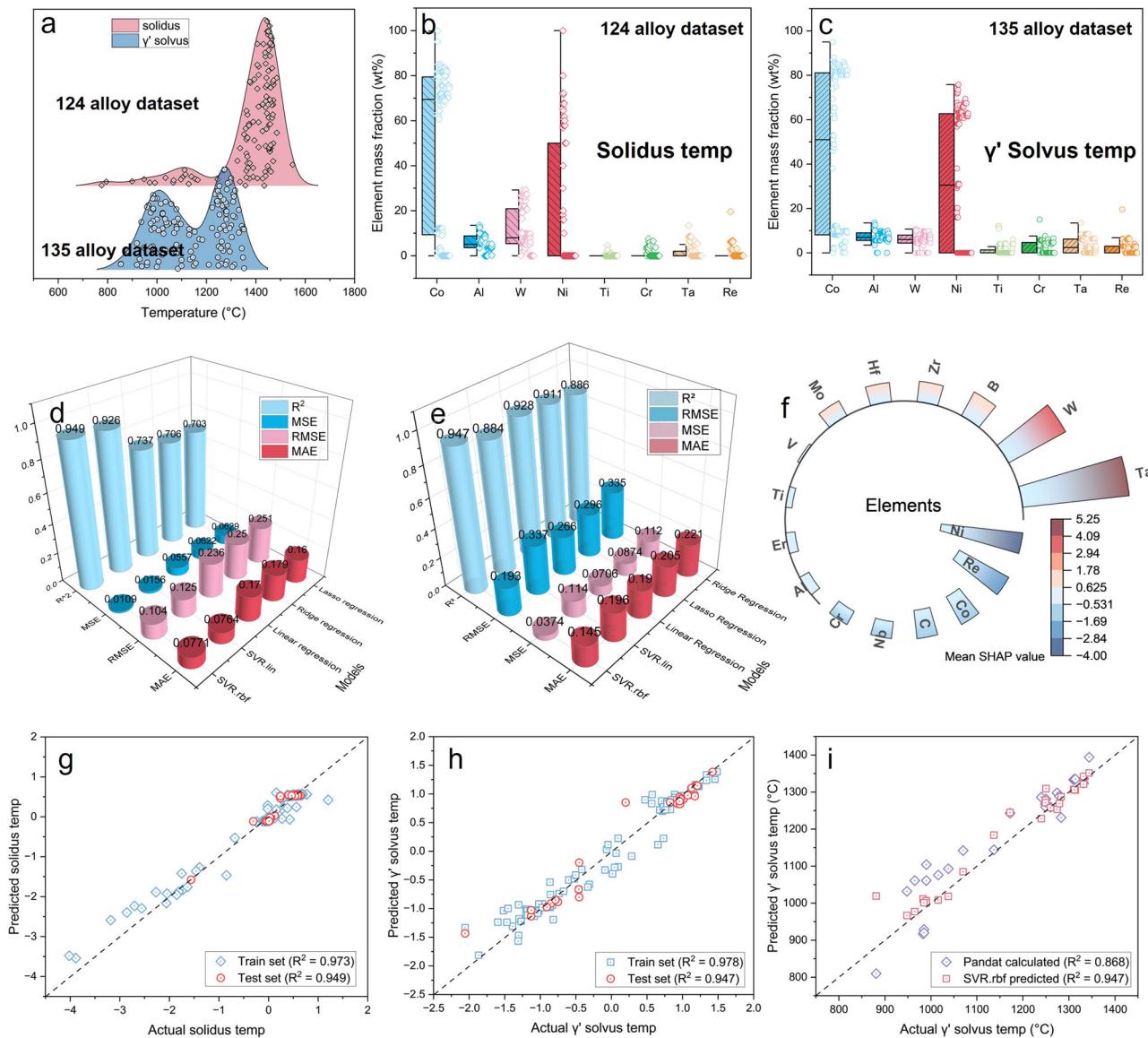


Fig. 4 | Comprehensive analysis of the machine learning-based prediction model for solidus and γ' solvus temperatures in single-crystal superalloys. a Distribution of experimentally measured solidus and γ' solvus temperatures across two datasets (124 and 135 alloys), highlighting their distinct thermal ranges. **b, c** Elemental composition distributions of the datasets. **d, e** Performance comparison of five regression models (linear regression, random forest, XGBoost, SVR.linear, and

SVR.rbf) on test sets, evaluated by R^2 , MSE, RMSE, and MAE; SVR.rbf achieves the highest predictive accuracy. **f** SHAP summary plot illustrating the contribution of each alloying element to γ' solvus temperature prediction, with Ta and W identified as the most influential. **g, h** Predicted vs. actual values for solidus and γ' solvus temperatures. **i** Comparison between SVR.rbf predictions and commercial thermodynamic software (Pandat) for γ' solvus temperature.

temperature and an R^2 of 0.947 for the γ' solvus temperature, significantly outperforming other models. To further enhance model credibility, we employed SHAP (SHapley Additive exPlanations) for interpretability analysis (Fig. 4f). It should be noted that SHAP values reflect statistical patterns learned by the model under the current training data distribution, which is inherently shaped by historical alloy design practices. The positive SHAP contributions of W and Ta primarily arise from their frequent co-occurrence with Re in high-performance alloys, rather than from a direct physical effect on γ' solvus temperature. Conversely, the apparent negative SHAP values for Al and Ti do not imply that these elements suppress the γ' solvus; instead, they stem from engineering constraints that actively limit their combined content to avoid excessive γ' volume fractions and mitigate TCP phase formation. This highlights a key limitation of purely data-driven approaches: the model captures how alloys have been designed, a reflection of human decision-making and historical precedent, rather

than the underlying physical laws governing how individual elements influence material properties. Scatter plots of actual versus predicted values (Fig. 4g, h) further validate the high accuracy of the model. Notably, we compared the SVR.rbf model with predictions from the commercial thermodynamic software Pandat (Fig. 4i). Using the Pan-Ni2022_TH + MB database from Pandat_2022 for γ' solvus temperature prediction on the same test set, Pandat achieved an R^2 of 0.868, whereas the SVR.rbf model reached 0.947, outperforming traditional thermodynamic methods. This demonstrates that machine learning models trained on high-quality experimental data can effectively capture complex composition-performance relationships and serve as reliable predictive tools for high-throughput alloy design.

Alloy design workflow

Building upon a structured materials database containing 135 experimentally measured γ' solvus temperatures and 124 solidus temperatures, we

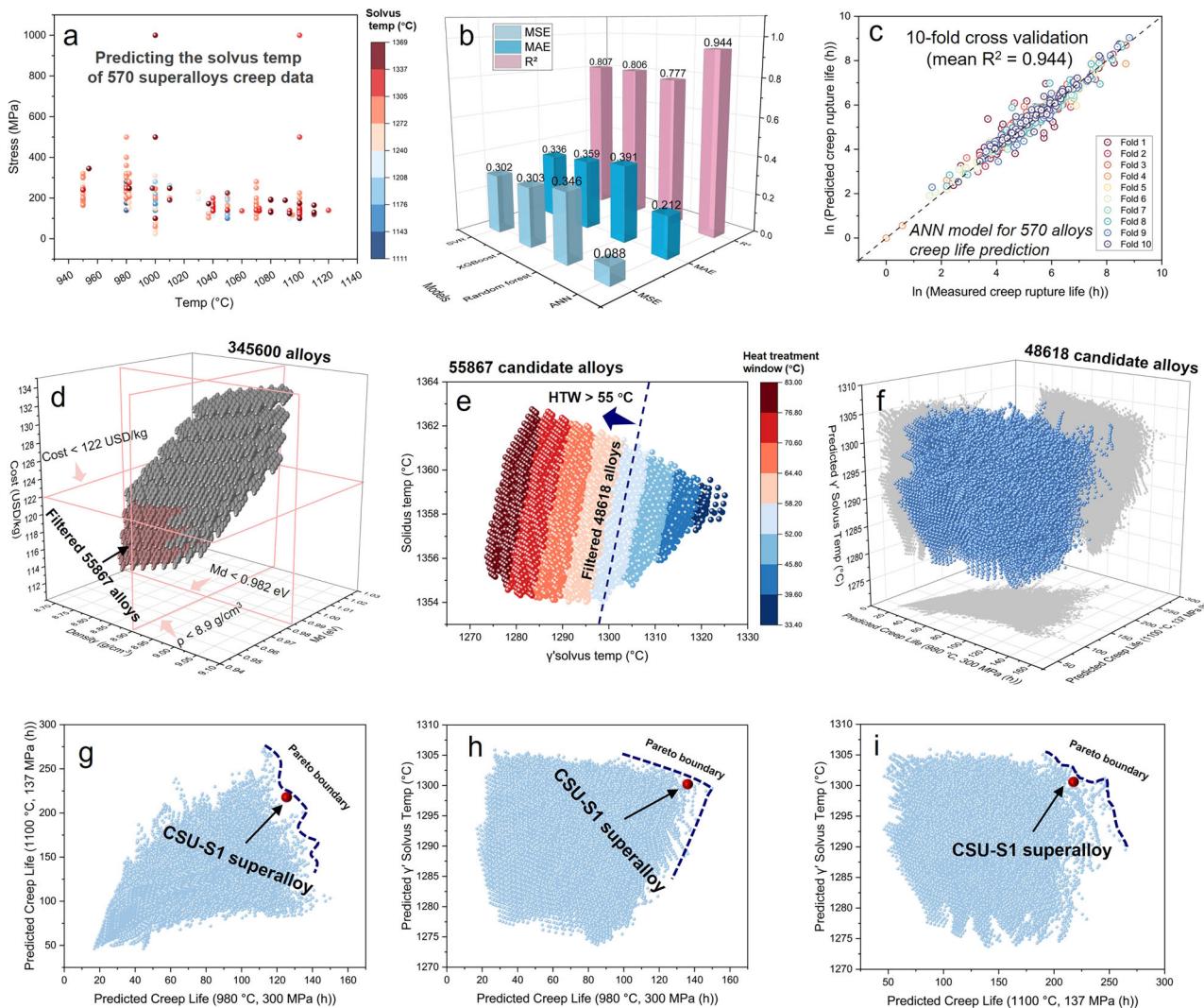


Fig. 5 | Data-driven high-throughput design and optimization of single-crystal superalloys. **a** Predicted γ' solvus temperatures for 570 known superalloys with experimentally measured creep dat. **b** Performance comparison of five machine learning models (linear regression, random forest, XGBoost, SVR, and ANN) in predicting creep rupture life, evaluated by R^2 , MSE, MAE, and RMSE; the ANN model achieves the highest accuracy. **c** 10-fold cross-validation results showing strong correlation between predicted and measured creep lives (mean $R^2 = 0.944$), demonstrating robust generalization. **d** High-dimensional design space of 345,600

Ni-based alloys, with initial screening based on density, cost, and microstructural stability ($Md < 0.982$ eV), reducing candidates to 55,867. **e** Further filtering using the heat treatment window (HTW > 55 °C), defined as solidus minus γ' solvus temperature, narrows the pool to 48,618 alloys. **f** Distribution of candidate alloys in the γ' solvus–creep life space after HTW screening. **(g–i)** Pareto fronts identified under multiple service conditions, the newly designed CSU-S1 alloy is highlighted near the Pareto boundary, indicating an optimal trade-off between γ' solvus temperature and creep resistance.

implemented a data-driven, high-throughput alloy design strategy. First, the trained SVR.rbf model was applied to predict the γ' solvus temperatures of 570 single-crystal superalloys with known creep performance (Fig. 5a; Table S4). These alloys were tested under representative aero-engine service conditions, with creep temperatures ranging from 950 to 1120 °C and applied stresses spanning from tens to 1000 MPa. To establish a complete mapping among composition, processing, and performance, we developed an artificial neural network (ANN) that takes alloy composition, predicted γ' solvus temperature, test temperature, and applied stress as inputs to regressively predict creep life.

To ensure model reliability, we benchmarked five machine learning algorithms, linear regression (LR), random forest (RF), XGBoost, support vector regression (SVR), and ANN, on the same dataset. As shown in Fig. 5b, the ANN model consistently outperformed all others across all evaluation metrics, achieving an R^2 of 0.944 on the test set, with the lowest mean absolute error (MAE) and mean squared error (MSE). Further validation via 10-fold cross-validation confirmed excellent agreement

between predicted and experimentally measured creep lives (Fig. 5c, demonstrating the model's strong generalization capability and robustness. Leveraging this ANN model, we constructed a high-dimensional design space encompassing 345,600 Ni-based single-crystal superalloy compositions (Fig. 5d) and systematically screened candidates using multi-physics constraints. The following boundary conditions were imposed: (i) Density < 8.9 g/cm³ (to meet lightweighting requirements); (ii) A high γ' solvus temperature indicates that the alloy can retain an effective volume fraction of γ' strengthening precipitates under high-temperature service conditions (e.g., 1100 °C). However, this does not equate to overall microstructural stability. On the contrary, an excessively high γ' volume fraction often leads to elevated concentrations of elements such as Cr and Re in the γ matrix, thereby increasing the risk of topologically close-packed (TCP) phase precipitation. We use the microstructural stability parameter $Md < 0.982$ eV (where Md is an electronic-structure-based descriptor derived from the d-band center, commonly used for rapid assessment of topologically close-packed (TCP) phase

precipitation tendency in Ni-based superalloys²⁶. It should be emphasized that Md provides only a rough, empirical estimate of stability; more rigorous phase stability analysis can be achieved via CALPHAD thermodynamic modeling. However, given the scale of high-throughput screening involving hundreds of thousands of alloys, CALPHAD calculations were not feasible in the current workflow; (iii) Material cost < 122 USD/kg (to ensure economic viability).

This initial filtering reduced the candidate pool from 345,600 to 55,867 alloys (Fig. 5e). Subsequently, we introduced the heat treatment window (HTW), defined as the difference between solidus and γ' solvus temperatures, as a key criterion for process feasibility. An insufficiently wide HTW can lead to a limited solution heat-treatment window, increasing the risk of incipient melting or microsegregation. Using the SVR.rbf model to predict both solidus and γ' solvus temperatures for 55867 candidates, we enforced a hard constraint of $HTW > 55$ °C, further narrowing the candidate set to 48,618 alloys. The ANN model was then employed to predict creep life under multiple service conditions for these remaining candidates. Finally, we identified the Pareto front in the two-dimensional performance space of γ' solvus temperature versus creep life (Fig. 5g-i) and selected a new alloy, CSU-S1, exhibiting an optimal balance of high-temperature strength and creep resistance.

The composition of CSU-S1 (wt%) is Al 5.6, Co 10.8, Cr 4.0, Mo 1.7, Re 3.1, Ta 6.6, Nb 0.05, W 8.0, balance Ni. Its predicted γ' solvus temperature reaches 1301.3 °C. Under representative service conditions, such as 980 °C/300 MPa and 1100 °C/137 MPa, its predicted creep lives lie close to the Pareto front (Fig. 5h, i), underscoring its exceptional high-temperature performance.

Alloy fabrication and property validation

To validate the accuracy of the alloy design and the reliability of the performance prediction model, [001]-oriented CSU-S1 single-crystal bars were fabricated using the spiral grain selector technique. The as-cast microstructure (Fig. 6a) exhibits a typical dendritic morphology, with pronounced solute enrichment in the inter-dendritic regions, indicating significant segregation during solidification. To eliminate this segregation and achieve compositional homogenization, a multi-stage solution heat treatment was employed: 1300 °C for 2 h, followed by 1310 °C for 4 h, and finally 1313 °C for 12 h, to promote atomic diffusion. After this treatment, the dendritic structure was eliminated, and the matrix became homogeneous (Fig. 6b), confirming the effectiveness of the solutioning protocol. Subsequently, a two-step aging treatment was applied to precipitate a regular γ' microstructure (Fig. 6c). The γ' precipitates are cuboidal in shape, with a bimodal size distribution (Fig. 6e) and an average size of 0.33 ± 0.09 μm . This dual-scale morphology is beneficial for achieving an optimal balance between strength and creep resistance.

To experimentally determine the γ' solvus and solidus temperatures, differential scanning calorimetry (DSC) was performed on the solution-treated sample (Fig. 6d). During heating, an endothermic peak appears at 1298.2 °C, corresponding to the γ' solvus temperature. A second, higher-temperature endothermic peak is observed at 1353.3 °C, attributed to the onset of melting and thus identified as the solidus temperature. Notably, the experimentally measured γ' solvus temperature shows excellent agreement with the value predicted by the SVR.rbf model (1301.3 °C; see Fig. 5h, i), thereby validating the high predictive accuracy of the machine learning framework.

Atomic-scale compositional analysis of the CSU-S1 single-crystal alloy was conducted using HAADF-STEM in combination with EDS. The HAADF-STEM image (Fig. 6f) and elemental maps (Fig. 6g) reveal Al enrichment in the γ' phase, consistent with its role as the main constituent of Ni_3Al , whereas Co and Cr preferentially partition to the FCC γ matrix. FFT analysis (Fig. 6h) further confirms the ordered L1_2 structure of γ' , as evidenced by superlattice reflections at the (100) and (110) extinction positions of the FCC lattice. In contrast, the γ matrix exhibits a random substitutional solid solution, with Co and Cr occupying Ni sites, and thus shows no additional superlattice reflections.

High temperature creep behavior and deformation mechanisms

Figure 7a presents the engineering displacement versus time curves for CSU-S1 alloy under various temperature-stress combinations. At 1100 °C and 137 MPa, the alloy exhibits a low creep rate over the first 220 h, underscoring its exceptional creep resistance. Under the intermediate condition of 980 °C and 300 MPa, CSU-S1 also demonstrates a long creep life and markedly higher ductility. The origin of this behavior is discussed below. It also presents a comparison between the predicted creep life of the CSU-S1 alloy by the ANN model and the actual measured values. The results show that the ANN model, which incorporates γ' solvus temperature as an input feature, demonstrates excellent predictive capability with errors of only about 20 h under various creep conditions. This confirms the effectiveness of ML for predicting alloy performance. Figure 7b compares the Re content and cost of CSU-S1 with those of several single-crystal superalloys of different generations, as shown in Table 1. Re is a scarce, high-value strategic metal. The Re content of CSU-S1 is 3.1 wt %, substantially lower than that of third-generation alloys such as DD9 (4.5 wt %), conferring a significant cost advantage. Figure 7b shows that the cost of CSU-S1 is 121 USD kg^{-1} , well below those of DD9 (162 USD kg^{-1} , 3rd-generation) and SC180 (143 USD kg^{-1} , 2nd-generation), further confirming its economic superiority.

Figure 7c and d present the stress distributions of CSU-S1 and reference alloys on the Larson–Miller parameter plot. CSU-S1, marked by red stars, sustains higher stress at lower Larson–Miller parameters and falls within the band of third-generation alloys, clearly outperforming second-generation Ni-base single-crystal superalloys. Under creep conditions at 1100 °C, CSU-S1 achieved a Larson–Miller parameter comparable to third-generation DD9 single-crystal alloys, outperforming the DD33 and René N6 alloy. This demonstrates that CSU-S1 delivers exceptional high-temperature strength while keeping both Re content and cost low. Collectively, CSU-S1 exhibits superior creep resistance, reduced Re usage and expense, and outstanding strength under high temperature, positioning it as a high-performance alloy with broad application potential. To further compare the high-temperature creep resistance of different alloys, key physical parameters of CSU-S1, CMSX-4 and DD9 were calculated at 980 °C and 1100 °C with Pandat_2022. Under low-stress creep (137 MPa) at high temperature (1100 °C), dislocations in the γ matrix bypass γ' precipitates via diffusion-assisted climb, a process that dominates diffusion-controlled deformation^{27,28}. Increasing γ' volume fraction and γ -matrix solid-solution strengthening hinder this climb, thus enhancing creep strength. At 1100 °C, CSU-S1 retains a high γ' fraction (54.21%) comparable to DD9 (52.14%). Furthermore, the solid-solution strengthening index (I_{SS})²⁹ of CSU-S1 is 12.88 at%, which is significantly higher than that of CMSX-4 at 9.59 at%, explaining its third-generation-level creep resistance. At lower temperatures, creep mechanisms become more complex, involving dislocation cutting of γ' to form anti-phase boundaries, stacking faults and micro-twins, all governed by antiphase boundary (APB) energy and stacking fault energy (SFE). Higher APB and SFE values suppress dislocation glide and cutting, thereby improving creep resistance. It is worth noting that at 980 °C, the APB energy of CSU-S1 reaches 0.223 J/m² and the SFE is 0.208 J/m², both higher than those of CMSX-4. Therefore, it demonstrated better creep resistance than the second-generation alloys.

To elucidate the creep deformation mechanisms of the CSU-S1 alloy at 980 °C, a systematic transmission electron microscopy (TEM) analysis was performed on crept-ruptured specimens, with representative results shown in Fig. 8. Low-magnification HAADF imaging (Fig. 8a) reveals that the γ' (L1_2) precipitates have coarsened significantly and aligned perpendicular to the stress axis, forming a characteristic transverse rafted microstructure. This morphological evolution arises from the interplay between anisotropic γ/γ' interfacial energy and stress-driven diffusion at elevated temperatures, underscoring the critical role of atomic diffusion during creep. The rafted structure acts as a geometric barrier, effectively impeding direct dislocation shearing of the γ' phase.

Higher-magnification bright-field imaging (Fig. 8b) shows that a high density of dislocations accumulates at the γ/γ' interfaces, forming tangled

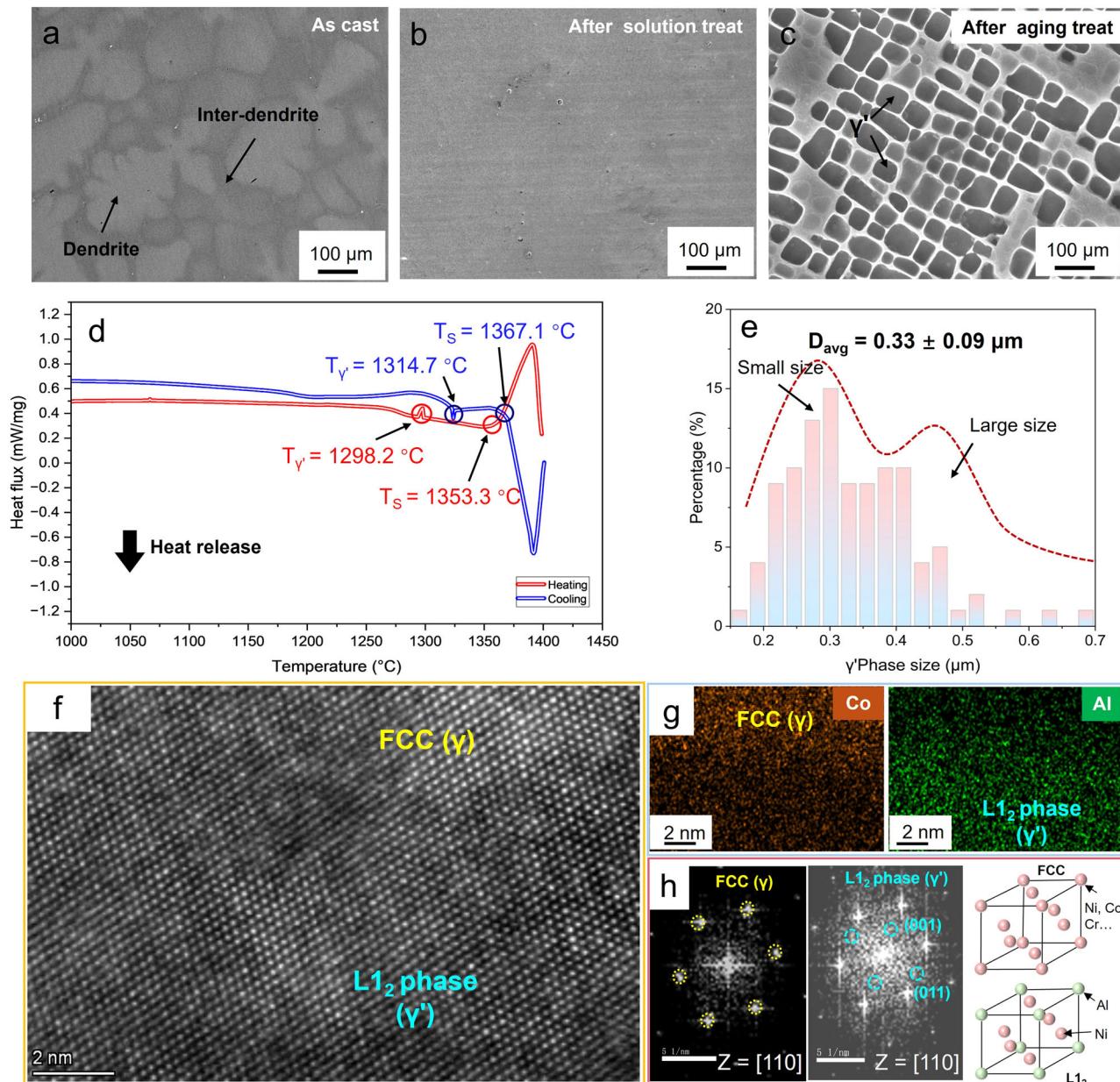


Fig. 6 | Microstructural evolution and phase-transformation characteristics of the CSU-S1 single-crystal superalloy. a As-cast: typical dendritic structure with pronounced inter-dendritic segregation. **b** After homogenization: multi-step solution treatment ($1300\text{ }^{\circ}\text{C}/2\text{ h} + 1310\text{ }^{\circ}\text{C}/4\text{ h} + 1313\text{ }^{\circ}\text{C}/12\text{ h}$) completely eliminates dendrites and yields a uniform γ matrix. **c** After two-step aging: regular arrays of

cuboidal γ' precipitates within the γ matrix. **d** DSC heating trace: endothermic peak at $1298.2\text{ }^{\circ}\text{C}$ corresponds to the γ' solvus; the second peak at $1353.3\text{ }^{\circ}\text{C}$ marks the solidus. The measured solvus agrees within $3\text{ }^{\circ}\text{C}$ with the SVR.rbf prediction ($1301.3\text{ }^{\circ}\text{C}$). **e** γ' size distribution. **f** Atomic-resolution HAADF-STEM image. **g** EDS elemental map. **h** FFT pattern.

networks. This indicates that dislocation motion within the γ channels is strongly hindered by the rafted γ' phase. Nevertheless, a limited number of dislocations are observed to cut through the γ' precipitates (Fig. 8c), suggesting that localized shearing can still occur under high-stress concentrations. High-resolution TEM imaging further reveals the presence of localized planar defects within the γ' phase (Fig. 8d), identified as super-lattice extrinsic stacking faults (SESFs)³⁰. Atomic-resolution energy-dispersive X-ray spectroscopy (EDS) mapping (Fig. 8(e)) shows pronounced enrichment of Cr and Co at the SESF cores³¹, accompanied by a relative depletion of Al. This component segregation is believed to occur after the formation of SESF, which can reduce the local stacking fault energy and thus stabilize it^{32,33}. Notably, the SESFs appear as isolated, non-continuous defects, without forming extended shear bands or through-thickness deformation zones. This observation, combined with thermodynamic

calculations indicating a high anti-phase boundary (APB) energy of 0.223 J/m^2 ^{34–36}, implies a substantial energy barrier against conventional dislocation shearing of the γ' phase. Consequently, under the high-temperature creep condition of $980\text{ }^{\circ}\text{C}$ (approximately 0.73 T_s), deformation is primarily accommodated by dislocation glide and thermally activated climb within the γ channels, a process further reinforced by the transverse rafting of γ' precipitates. The formation of isolated SESFs serves as a secondary, localized mechanism that helps relieve stress concentrations and enhances deformation homogeneity.

In summary, CSU-S1 exhibits coexisting creep deformation behavior at $980\text{ }^{\circ}\text{C}$, dominated by interface-pinned dislocation motion with limited γ' shearing via SESFs as a complementary pathway. This synergistic mechanism enables the alloy to maintain high creep resistance while retaining good creep ductility.

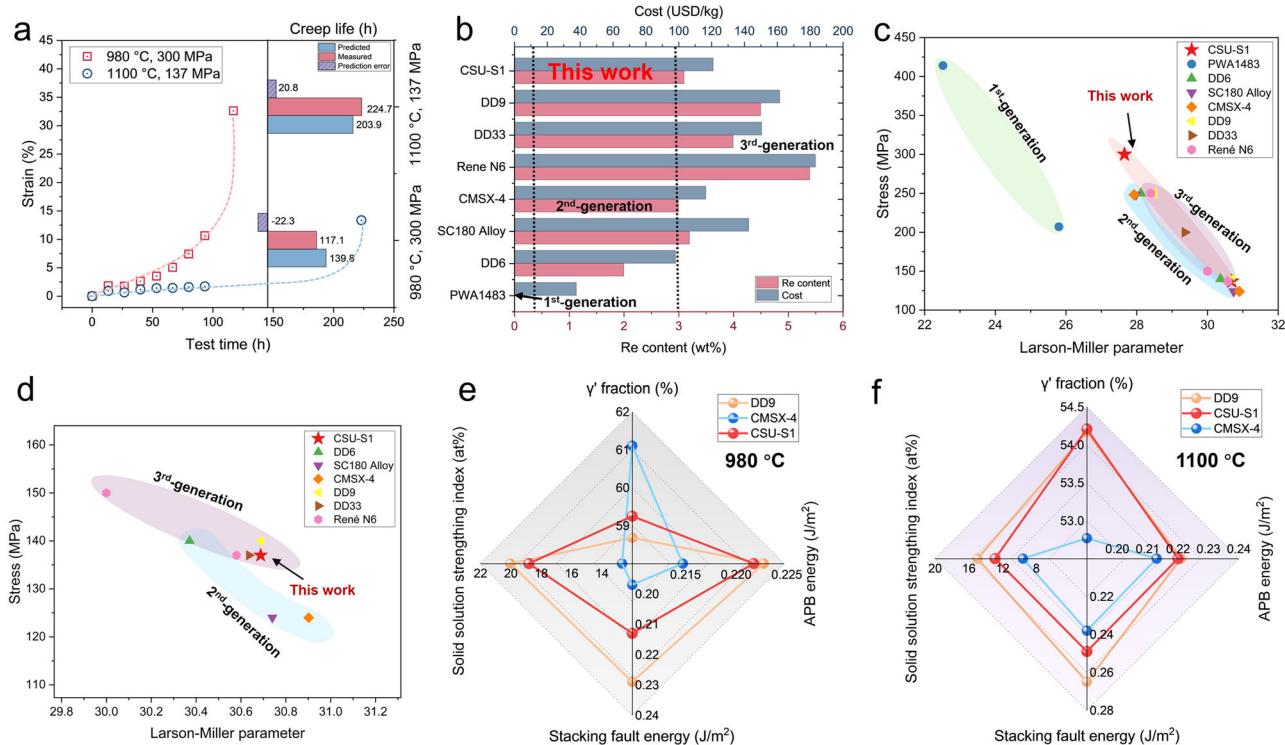


Fig. 7 | Comparison of the CSU-S1 alloy with other superalloys. a Creep curves of CSU-S1 under two different stress conditions: 980 °C, 300 MPa and 1100 °C, 137 MPa. **b** Re content and cost comparison among CSU-S1, CMSX-4, DD9, DD6, PWA1483, Rene N6, and SC180 Alloy. **c, d** Stress and Larson-Miller parameter plot for various alloys^{37–42}. **e, f** Thermodynamically derived high-temperature physical parameters for the alloys.

Table 1 | Chemical composition (wt%) of different single-crystal superalloys

Alloy	Al	Co	Cr	W	Mo	Re	Ta	Ti	Nb	Ni	Hf
CSU-S1	5.6	10.8	4	8	1.7	3.1	6.6	0.6	0.05	59.55	0
DD9	5.6	7	3.5	6.5	2	4.5	7.5	0	0.5	63.3	0.1
DD33	6	9	3.5	6	1.5	4	8	0.2	0	61.7	0.1
Rene N6	5.75	12.5	4.2	6	1.4	5.4	7.2	0	0	57.4	0.15
CMSX-4	5.6	9	6.5	6	0.6	3	6.5	1	0	61.8	0
DD6	5.6	9	4.3	8	2	2	7.5	0	0.5	61.1	0
SC180 Alloy	5.81	10.5	5.91	4.51	1.49	3.42	9.8	1	0	57.46	0.1
PWA1483	3.63	8.7	11.98	3.83	1.81	0	5.09	3.99	0	60.97	0.005

Discussion

In this work, we established a closed-loop, data-driven framework for the intelligent design of single-crystal superalloys by integrating natural language processing (NLP) and machine learning (ML).

First, the domain-adapted NLP model Superalloy-MatSciBERT was employed to efficiently extract critical thermodynamic data, specifically γ' solvus temperatures, from vast amounts of unstructured scientific literature. This approach effectively addresses the persistent data scarcity bottleneck in high-performance alloy development.

Building on this curated dataset, we developed high-accuracy predictive models: an SVR with RBF kernel (SVR.rbf) for γ' solvus temperature and an artificial neural network (ANN) for creep life. Guided by these models and incorporating multi-dimensional engineering constraints, including cost, density, phase stability, and processability, we screened hundreds of thousands of candidate compositions and successfully identified a new alloy, CSU-S1.

Experimental validation confirmed that CSU-S1 achieves creep performance comparable to third-generation single-crystal superalloys while significantly reducing rhenium content and material cost, demonstrating

strong potential for practical applications. Comprehensive microstructural characterization reveals that its superior high-temperature performance stems from two complementary mechanisms: (i) strong resistance to dislocation motion provided by transverse γ' rafting, (ii) localized plasticity accommodation enabled by isolated superlattice extrinsic stacking faults (SESFs), stabilized through Cr/Co segregation.

Beyond this specific alloy system, our study validates the feasibility and efficiency of an NLP- and ML-driven materials design paradigm. It offers a generalizable framework that transforms latent knowledge embedded in scientific literature into a computable, predictable, and optimizable engine for innovation, accelerating the intelligent development of advanced engineering materials across diverse domains.

Methods

Corpus acquisition and pre-processing

We collected journal articles (2010–2023) from Elsevier, Springer, and Nature via the CrossRef REST API using the keyword “single-[]crystal.*-superalloy” and publisher-specific DOI prefixes. Patents were retrieved from Patent-Star using a Selenium-controlled headless Chrome driver. All

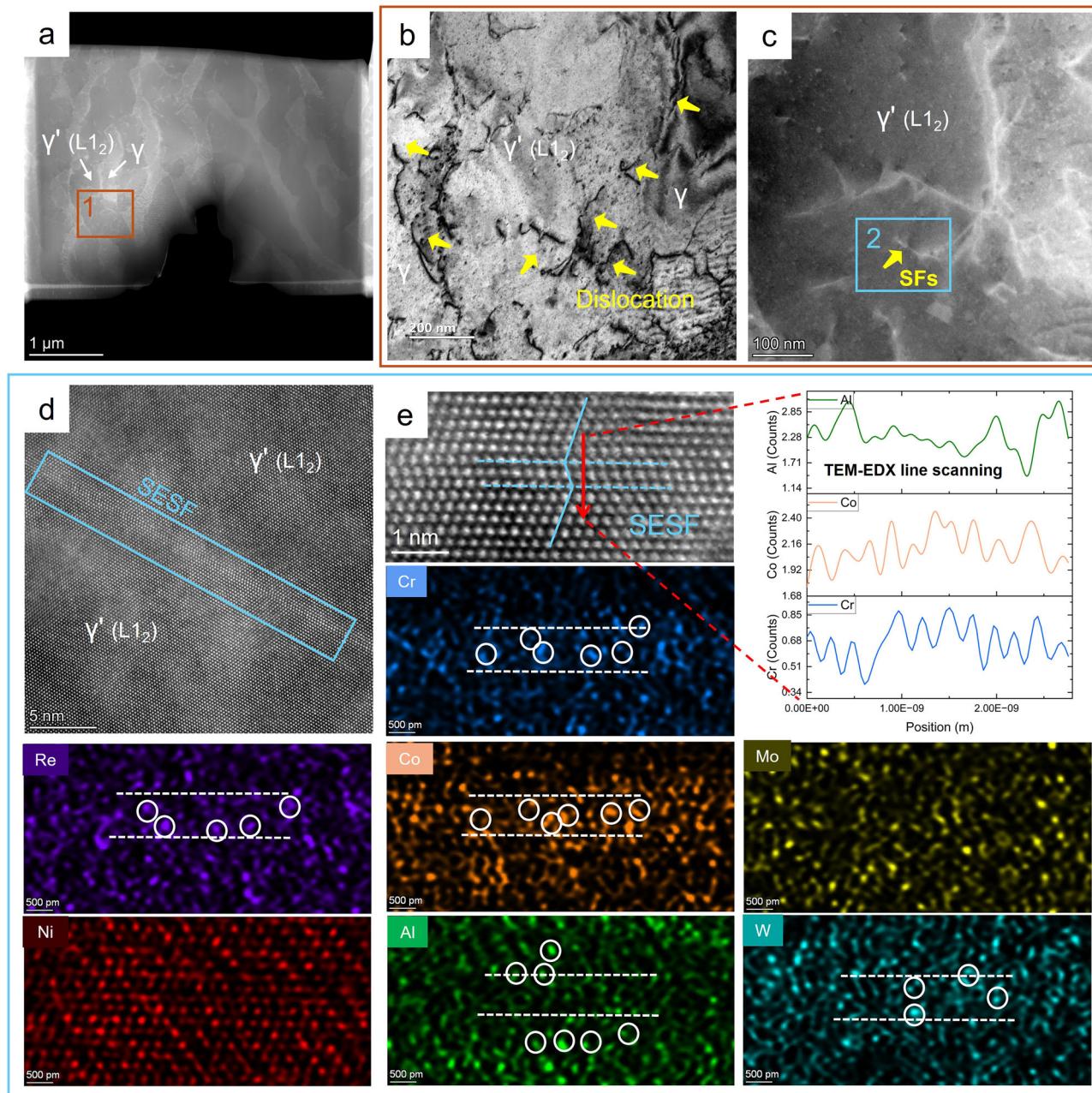


Fig. 8 | Microstructural characterization of the CSU-S1 alloy after creep testing at 980 °C/300 MPa, revealing deformation mechanisms. **a** Low-magnification HAADF-STEM image showing transverse rafting of γ' ($L1_2$) precipitates perpendicular to the stress axis. **b** Bright-field TEM image illustrating a high density of dislocations accumulated at γ/γ' interfaces and forming tangled networks.

c Enlarged view highlighting dislocations cutting through γ' precipitates (yellow arrows), suggesting localized shearing under high-stress regions. **d** High-resolution TEM image revealing superlattice extrinsic stacking faults (SESFs) within γ' phase. **e** Atomic-scale EDS elemental mapping and line scan across an SESF: Cr and Co are enriched, while Al is depleted at the fault core.

documents were converted to plain text, with HTML tags, equations, and reference lists removed via regular expressions. Text was sentence-tokenized using NLTK's Punkt tokenizer, enhanced with a domain-specific abbreviation dictionary, and temperature units were normalized to °C. Three supervised datasets, covering sentence classification, named entity recognition, and relation extraction, were independently annotated by two researchers and cross-validated. The data were stratified 6:2:2 into train/validation/test splits, with all sentences from a given paper assigned to a single split to prevent information leakage.

Superalloy-MatSciBERT framework

Built on a 12-layer, 768-d MatSciBERT backbone, the framework was fine-tuned for three tasks: sentence classification feeds the [CLS] vector through

Dense(50, ReLU) → Dense(2, Softmax) and uses Focal Loss ($\gamma=2$) to counter the 1:6 imbalance; named-entity recognition appends a BiLSTM(128) + CRF layer after MatSciBERT to capture label dependencies; relation extraction inputs entity-marked sentences into MatSciBERT → Dropout(0.2) → Dense(2, Sigmoid). All heads were trained with AdamW for 34, 100, and 103 epochs respectively (early stopping patience = 5, max length 256).

γ' solvus temperature prediction model

The 135 manually verified “25-dimensional elemental mass fraction– γ' solvus temperature” records were split 8: 2 into training and test sets and evaluated with five-fold cross-validation across six

models: linear, ridge, lasso, SVR-linear, SVR-RBF, and GBR. Grid search yielded the optimal SVR-RBF ($C = 1000$, $\gamma = 0.01$).

ANN creep life prediction model

Creep life prediction was cast as a multivariate regression task: inputs are alloy elemental mass fractions plus the predicted γ' solvus temperature, and the output is creep life (h). Owing to the long-tailed distribution of life, the raw values were log-transformed, and all features were column-wise Z-score normalized. 10-fold cross-validation was employed: the dataset was randomly stratified into ten equal parts, with nine folds used for training and one for validation in each iteration, repeated ten times to ensure robust evaluation. The network is a four-layer fully connected feed-forward ANN: input layer of 14 dimensions (creep test temperature, stress, 11 elemental mass fractions, and predicted γ' solvus temperature); first hidden layer—14 ReLU units with L_2 regularization 0.001; second—64 ReLU units; third—16 ReLU units; output layer—one linear unit for creep life. During cross-validation, MSE, MAE, and R2 were recorded for each fold; final metrics are the 10-fold averages.

Alloy composition design workflow

We generated a composition design space for single-crystal superalloys. The Al content ranges from 5.6 to 6 wt%, with a step size of 0.1 wt%; Co ranges from 8 to 12 wt%, with a step size of 0.2 wt%; Cr ranges from 4 to 6 wt%, with a step size of 0.2 wt%; W ranges from 6 to 10 wt%, with a step size of 0.2 wt%; Mo ranges from 0 to 3 wt%, with a step size of 0.1 wt%; Re ranges from 3 to 5 wt%, with a step size of 0.1 wt%; Ta ranges from 6 to 8 wt%, with a step size of 0.2 wt%; Ti ranges from 0 to 1 wt%, with a step size of 0.1 wt%; Nb ranges from 0 to 0.5 wt%, with a step size of 0.05 wt%. The final element, Ni, is the balancing element, and its wt% is calculated as 100 minus the sum of the other elements. A total of 345,600 alloy samples were generated.

The density of nickel-based superalloys was calculated using a modified version of the Caron model. The performance of the modified model is compared with that of the original Caron model as shown in Fig. S1. It can be observed that, at lower density ranges, for example, for DD3, SC16, and CMSX-6 alloys, the Caron density model tends to underestimate the alloy density. The updated density model exhibits excellent predictive capability over a broader range of densities.

The updated density modeling formula is as follows:

$$\rho(\text{g/cm}^3) = 8.939 - 0.0014c_{\text{Co}} - 0.0197c_{\text{Cr}} + 0.0125c_{\text{Mo}} + 0.0436c_{\text{W}} + 0.0508c_{\text{Re}} + 0.0194c_{\text{Ru}} - 0.1281c_{\text{Al}} - 0.0498c_{\text{Ti}} + 0.0409c_{\text{Ta}}$$

where c_i denotes the mass percentage of alloying element i .

Superalloys are materials that can operate stably at high temperatures over extended periods and are widely used in high-temperature components such as aeroengines and gas turbines. Therefore, their microstructural stability needs to be carefully assessed. In this study, the New PHACOMP method was employed to predict the microstructural stability of superalloys. The specific calculation method is as follows:

$$\bar{Md} = \sum_i^n x_i M_{di}$$

In the formula, x_i represents the atomic fraction of alloying element i , while M_{di} represents the energy level of the d-orbitals for element i . The Md values for common elements in nickel-based superalloys are summarized in Table S5.

The cost of the alloy was calculated based on the unit price of each element in the composition. The prices of the individual elements are as follows: Al (2.89 USD/kg), Co (24.66 USD/kg), Cr (10.28 USD/kg), W (53.04 USD/kg), Mo (62.97 USD/kg), Re (2,464.79 USD/kg), Ta (405.21 USD/kg), Ti (5.63 USD/kg), Nb (94.93 USD/kg), Ni (16.90 USD/kg), and Hf (3,239.44 USD/kg). To calculate the alloy cost per kilogram, the mass fraction of each element in the alloy was multiplied by its corresponding unit

price in USD. The total cost of the alloy was obtained by summing the costs of all individual elements. This approach provides an accurate calculation of the alloy's cost based on its elemental composition.

Experiments

Raw materials with 99.9% purity were weighed according to the nominal composition. A master ingot was first melted in a dedicated VIM-1700 medium-frequency vacuum induction furnace. Single-crystal bars oriented along [001] were then produced by the spiral grain-selection technique. All subsequent heat-treatment steps were carried out in an MITR-1700X-8L high-temperature box furnace. Microstructures were examined in a Tescan-Mira4 scanning electron microscope. Transmission electron specimens were prepared with a TESCAN-AMBER focused-ion-beam (FIB) scanning electron microscope. High-resolution atomic imaging and energy-dispersive X-ray spectroscopy were performed in a double-corrected Spectra-300 TEM operated at 300 kV. High-temperature creep tests were conducted on an RWS100 electronic creep testing machine in accordance with ASTM standards at 980 °C–300 MPa and 1100 °C–137 MPa.

Data availability

The data that support the findings of this study are available in Tables S1 and S2. We have fully open-sourced the entire project, including all code, data processing pipelines, model training scripts, and dependency environment, and made it publicly available in the GitHub repository: <https://github.com/Yvonnehin/Superalloy-MatLitMiner>.

Received: 27 August 2025; Accepted: 29 November 2025;

Published online: 19 December 2025

References

- Long, H., Mao, S., Liu, Y., Zhang, Z. & Han, X. Microstructural and compositional design of Ni-based single crystalline superalloys—A review. *J. Alloy. Compd.* **743**, 203–220 (2018).
- Reed, R. C. *The superalloys: fundamentals and applications*. (Cambridge University Press, 2008).
- Darolia, R. Development of strong, oxidation and corrosion resistant nickel-based superalloys: critical review of challenges, progress and prospects. *Int. Mater. Rev.* **64**, 355–380 (2019).
- Pollock, T. M. Alloy design for aircraft engines. *Nat. Mater.* **15**, 809–815 (2016).
- Pollock, T. M. & Tin, S. Nickel-based superalloys for advanced turbine engines: chemistry, microstructure and properties. *J. Propuls. power* **22**, 361–374 (2006).
- Agrawal, A. & Choudhary, A. Perspective: Materials informatics and big data: Realization of the “fourth paradigm” of science in materials science. *APL Mater.* **4**, 053208 (2016).
- Guo, K. et al. Magnetic performance-oriented composition design of Sm-Co-based alloys by machine learning and experimental studies. *Comput. Mater. Sci.* **205**, 111232 (2022).
- Liu, Y. et al. Machine learning-enabled repurposing and design of antifouling polymer brushes. *Chem. Eng. J.* **420**, 129872 (2021).
- Ruan, J. et al. Accelerated design of novel W-free high-strength Co-base superalloys with extremely wide γ/γ' region by machine learning and CALPHAD methods. *Acta Mater.* **186**, 425–433 (2020).
- Liu, Y. et al. Predicting creep rupture life of Ni-based single crystal superalloys using divide-and-conquer approach based machine learning. *Acta Mater.* **195**, 454–467 (2020).
- Yao, J. et al. Machine learning guided insights into the effects of Nb/Ta and Ti/Ta ratios on microstructure and creep rupture life in Nickel-based single-crystal superalloys. *Metals Mater. Int.* <https://doi.org/10.1007/s12540-025-01899-7> (2025).
- Yao, J. et al. Integrating machine learning and thermodynamic descriptors for enhanced Ni-based single crystal superalloys creep life prediction and alloy design. *Metals Mater. Int.* <https://doi.org/10.1007/s12540-025-01906-x> (2025).

13. Liu, F. et al. High-throughput method-accelerated design of Ni-based superalloys. *Adv. Funct. Mater.* **32**, 2109367 (2022).
14. Bhadeshia, H., MacKay, D. & Svensson, L.-E. Impact toughness of C-Mn steel arc welds—Bayesian neural network analysis. *Mater. Sci. Technol.* **11**, 1046–1051 (1995).
15. Conduit, B. D., Jones, N. G., Stone, H. J. & Conduit, G. J. Design of a nickel-base superalloy using a neural network. *Mater. Des.* **131**, 358–365 (2017).
16. Menou, E. et al. Evolutionary design of strong and stable high entropy alloys using multi-objective optimisation based on physical models, statistics and thermodynamics. *Mater. Des.* **143**, 185–195 (2018).
17. Wang, W. et al. Automated pipeline for superalloy data by text mining. *npj Comput. Mater.* **8**, 9 (2022).
18. Court, C. J. & Cole, J. M. Auto-generated materials database of Curie and Néel temperatures via semi-supervised relationship extraction. *Sci. Data* **5**, 1–12 (2018).
19. Olivetti, E. A. et al. Data-driven materials research enabled by natural language processing and information extraction. *Appl. Phys. Rev.* **7**, 041317 (2020).
20. Rettig, R. et al. Development of a low-density rhenium-free single crystal nickel-based superalloy by application of numerical multi-criteria optimization using thermodynamic calculations. *Superalloys 2016*, 35–44 (2016).
21. Murakumo, T., Kobayashi, T., Koizumi, Y. & Harada, H. Creep behaviour of Ni-base single-crystal superalloys with various γ' volume fraction. *Acta Mater.* **52**, 3737–3744 (2004).
22. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. North American Chapter of the Association for Computational Linguistics. 4171–4186 (2019).
23. Beltagy, I., Lo, K. & Cohan, A. SciBERT: A Pretrained Language Model for Scientific Text. Association for Computational Linguistics. 3615–3620 (2019).
24. Gupta, T., Zaki, M., Krishnan, N. M. A. & Mausam. MatSciBERT: A materials domain language model for text mining and information extraction. *npj Comput. Mater.* **8**, 102 (2022).
25. Jang, B., Kim, I. & Kim, J. W. Word2vec convolutional neural networks for classification of news articles and tweets. *PLoS one* **14**, e0220976 (2019).
26. Morinaga, M., Yukawa, N., Adachi, H. & Ezaki, H. New Phacomp and its applications to alloy design. *Superalloys 1984*, 523–532 (1984).
27. Koizumi, Y. et al. in *Proceedings of the International Gas Turbine Congress, Tokyo*.
28. Zhang, J., Murakumo, T., Harada, H., Koizumi, Y. & Kobayashi, T. Creep deformation mechanisms in some modern single-crystal superalloys. *Superalloys 2004*, 189–195 (2004).
29. Rettig, R., Ritter, N. C., Helmer, H. E., Neumeier, S. & Singer, R. F. Single-crystal nickel-based superalloys developed by numerical multi-criteria optimization techniques: design based on thermodynamic calculations and experimental validation. *Model. Simul. Mater. Sci. Eng.* **23**, 035004 (2015).
30. Li, X., Zhang, H., Jia, J., Liu, J. & Zhang, Y. A twinning mechanism: Direct observation of the transition from stacking faults to Microtwins. *Scr. Mater.* **267**, 116803 (2025).
31. Smith, T. M. et al. Segregation and η phase formation along stacking faults during creep at intermediate temperatures in a Ni-based superalloy. *Acta Mater.* **100**, 19–31 (2015).
32. Viswanathan, G. B. et al. Segregation at stacking faults within the γ' phase of two Ni-base superalloys following intermediate temperature creep. *Scr. Mater.* **94**, 5–8 (2015).
33. Yamashita, M. & Kakehi, K. Tension/compression asymmetry in yield and creep strengths of Ni-based superalloy with a high amount of tantalum. *Scr. Mater.* **55**, 139–142 (2006).
34. Vorontsov, V. A., Kovarik, L., Mills, M. & Rae, C. High-resolution electron microscopy of dislocation ribbons in a CMSX-4 superalloy single crystal. *Acta Mater.* **60**, 4866–4878 (2012).
35. Knowles, D. M. & Chen, Q. Superlattice stacking fault formation and twinning during creep in γ/γ' single crystal superalloy CMSX-4. *Mater. Sci. Eng.: A* **340**, 88–102 (2003).
36. Chen, Q. Z. & Knowles, D. M. Mechanism of $\langle 112 \rangle /3$ slip initiation and anisotropy of γ' phase in CMSX-4 during creep at 750°C and 750 MPa. *Mater. Sci. Eng.: A* **356**, 352–367 (2003).
37. Wilson, B. & Fuchs, G. The effect of composition, misfit, and heat treatment on the primary creep behavior of single crystal nickel base superalloys PWA 1480 and PWA 1484. *Superalloys 2008*, 149–158 (2008).
38. EARL, W. R., KEVIN, S. O. & HARA, R. N4: A first generation single crystal turbine airfoil alloy with improved oxidation resistance, low angle boundary strength and superior long time rupture strength [A]. *Kissinger RD, Deye DJ, Anton DL, et al. Superalloys 1996*, 1919–1925 (1996).
39. Wang, X. G., Li, J. R., Shi, Z. X. & Liu, S. Z. in *Materials Science Forum*. 549–558 (Trans Tech Publ).
40. Ma, A., Dye, D. & Reed, R. A model for the creep deformation behaviour of single-crystal superalloy CMSX-4. *Acta Mater.* **56**, 1657–1670 (2008).
41. Li, J. et al. A Low-cost second-generation single crystal superalloy DD6. *Superalloys 1*, 777–783 (2000).
42. Pandey, A. & Hemker, K. J. Temperature dependence of the anisotropy and creep in a single-crystal nickel superalloy. *JOM* **67**, 1617–1623 (2015).

Acknowledgements

This work is supported by the Aero Engine Corporation of China [Grant no. HFZL2022CXY029], Young Elite Scientists Sponsorship Program by CAST [2022QNRC001]. We acknowledge the High-Performance Computing Center of Central South University, Powder Metallurgy National Key Laboratory, Electron Microscopy Laboratory, and Project Supported by State Key Laboratory of Powder Metallurgy, Central South University, Changsha, China.

Author contributions

L.M.T., L.H., and F.L. conceived, supervised, and acquired funding for the project. J.Y. and Z.W. contributed equally to this work; they jointly developed the NLP/ML framework, designed the workflow, performed modeling and data analysis, and co-wrote the manuscript. J.C.W. synthesized the single-crystal alloys and conducted mechanical- and microstructure-characterization experiments. W.C.Y. and Y.X.C. constructed and validated the NLP pipeline, extracted literature data, and curated the datasets. W.F.L. carried out high-temperature creep testing and DSC measurements. J.H.W. and Y.X.Z. performed SEM, TEM, and FIB analyses. Y.W. and L.W. assisted in sample preparation and property characterization. Y.L. and L.M.T. coordinated the experimental program and supervised materials processing. L.H. and F.L. revised the manuscript with input from all authors. All authors discussed the results and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-025-01906-w>.

Correspondence and requests for materials should be addressed to Liming Tan, Lan Huang or Feng Liu.

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

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025