Research Article

Transfer learning-based prediction of high-temperature fatigue performance in Fe-based structural alloys with limited data

Qi Wang[1,#], Chunlei Shang[1,#], Hong-Hui Wu[1,2,3,*], Dexin Zhu[1], Shuize Wang[1,3], Junheng Gao[1,3], Haitao Zhao[1,3], Chaolei Zhang[1,3], Yuhe Huang[1,3], Jun Lu[1,3], Xinping Mao[1,3]

[1]Beijing Advanced Innovation Center for Materials Genome Engineering, Institute for Carbon Neutrality, University of Science and Technology Beijing, Beijing 100083, China.
[2]Institute of Materials Intelligent Technology, Liaoning Academy of Materials, Shenyang 110004, Liaoning, China.
[3]Institute of Steel Sustainable Technology, Liaoning Academy of Materials, Shenyang 110004, Liaoning, China.
[#]Authors contributed equally.

[*]**Correspondence to:** Prof. Hong-Hui Wu, Beijing Advanced Innovation Center for Materials Genome Engineering, Institute for Carbon Neutrality, University of Science and Technology Beijing, Beijing 100083, China. E-mail: wuhonghui@ustb.edu.cn

How to cite this article: Wang Q, Shang C, Wu HH, Zhu D, Wang S, Gao J, Zhao H, Zhang C, Huang Y, Lu J, Mao X. Transfer learning-based prediction of high-temperature fatigue performance in Fe-based structural alloys with limited data. *J Mater Inf* 2026;6:[Accept]. http://dx.doi.org/10.20517/jmi.2026.06

Received: 23 February 2026 | **Revised:** 19 May 2026 | **Accepted:** 22 May 2026

Abstract

High-temperature fatigue life serves as a crucial performance metric for assessing the structural integrity and service safety of structural materials under elevated-temperature conditions. However, conventional assessment approaches rely on high-temperature fatigue testing, which is typically time-consuming, costly, and experimentally challenging. Consequently, the scarcity of reliable high-temperature fatigue data poses a major obstacle to accurate fatigue life prediction and limits the effectiveness of conventional machine-learning models. In this study,

a transfer-learning framework is proposed to address this data limitation by leveraging data-rich room-temperature fatigue datasets. An optimal feature subset is used with a Gradient Boosting Decision Tree model, and limited high-temperature samples are progressively incorporated through incremental retraining to enable effective knowledge transfer across temperature domains. The results demonstrate significantly improved predictive accuracy and confirm the consistency of the dominant features across room and high temperatures. Overall, the proposed framework offers a practical strategy for materials design and high-temperature fatigue-life assessment in engineering applications.

Keywords: High-temperature fatigue, fatigue life prediction, transfer learning, gradient boosting decision tree, feature selection

INTRODUCTION

Fatigue failure represents a predominant concern in engineering, accounting for 80-90% of structural fractures in mechanical components [1]. Thus, the precise assessment and prediction of the fatigue behavior of metals have substantial scientific and engineering relevance. Fatigue life refers to the number of stress cycles a material can withstand before fatigue failure under cyclic loading [2] and is influenced by multiple factors. These factors include the magnitude of the applied stress, the material's intrinsic properties, and environmental conditions such as temperature [3-5]. While the staircase method [6] and similar tests are necessary to obtain fatigue life data, these experimental procedures are both costly and time-consuming. Furthermore, the evaluation of fatigue properties is complicated by the diversity of required test conditions, including test type, frequency, and loading parameters. These challenges are further exacerbated when fatigue behavior at elevated temperatures is considered, as high-temperature fatigue tests are particularly difficult to conduct and the corresponding experimental data are scarce [7,8].

To predict fatigue properties, a range of conventional models have been developed. These range from empirical formulations, such as the Coffin-Manson and Basquin equations [9], to physics-based approaches including the critical plane method [10], multiaxial fatigue damage models [11], energy-based methods [12], and continuum damage mechanics [13]. While valuable, these methods often rely on idealized assumptions or extensive data fitting. The complex, nonlinear interactions among factors such as manufacturing processes and service conditions are difficult to capture accurately, limiting predictive accuracy under practical uncertainties. Machine

learning (ML) has recently become a powerful tool for fatigue analysis, offering superior capabilities in processing complex datasets and uncovering nonlinear patterns, enabling more accurate predictions than conventional methods [14]. It has shown promising results in predicting fatigue life [15-17], fatigue crack growth rate [18,19], and fatigue strength [20-22]. However, a significant gap remains in accurately predicting high-temperature fatigue performance due to the scarcity of relevant data. This persistent challenge underscores the critical need for innovative approaches to achieve reliable predictions under such conditions.

As a machine learning paradigm, transfer learning (TL) leverages knowledge acquired from one task or domain to facilitate learning another related but distinct task or domain [23]. TL primarily aims to reduce the amount of training data required to develop a high-performing model in the target domain by leveraging pre-existing knowledge from a related source domain [24,25]. The process involves training a model on a source domain with sufficient data, then transferring and adapting this knowledge to a target domain where data acquisition is difficult. In fatigue prediction, TL has proven effective and demonstrated strong predictive performance. To enable accurate prediction of gear contact fatigue life, Li et al. [26] proposed a transfer learning-based strategy that leveraged data from other low-cost fatigue tests to enhance predictive performance. To enable accurate prediction of the high-cycle fatigue Stress-Number of cycles (S-N) curve of steel, Wei et al. [27] proposed a novel prediction framework that integrates long short-term memory networks with TL, demonstrating high accuracy in predicting S-N curves across various steel types. Similarly, Zhai et al. [28] proposed a TL-enhanced physics-informed neural network that transfers defect-fatigue knowledge from additive manufacturing to megacasting, enabling high-precision fatigue life predictions with limited data. Despite these advances, existing TL methods still face inherent limitations when applied to cross-temperature fatigue prediction. Many of these approaches assume significant similarity in features or distributions between the source and target domains, an assumption that may not hold under drastically different thermal conditions. Moreover, most current models lack mechanisms to quantify or adapt to temperature-induced physicochemical changes explicitly, undermining the reliability of transferred knowledge. These limitations, together with the scarcity of high-temperature data, constrain predictive accuracy and generalizability in practical high-temperature applications, underscoring the need for novel approaches to bridge the domain gap between room and elevated temperatures effectively.

To address this problem, a TL strategy is proposed that uses room-temperature fatigue data to

predict high-temperature fatigue life. Six machine learning models are trained on room-temperature fatigue data, and their performance is evaluated both with and without feature selection. Subsequently, after verifying feature consistency across temperature ranges, the optimal model and the critical feature subset are selected. The pre-trained model is then incrementally retrained on a subset of high-temperature data to predict high-temperature fatigue life. This strategy provides an effective and data-efficient framework for improving fatigue life prediction under high-temperature conditions.

METHODOLOGY

A TL framework is proposed to address the limited availability of high-temperature fatigue data by leveraging knowledge from extensive room-temperature datasets, as shown in Fig. 1. Initially, to establish a robust and efficient baseline, six ML models are evaluated using the full feature set from the room-temperature data. Subsequent feature selection identified an optimal subset of five features. A comparison between the optimal feature subset and the full feature set using the same models shows no significant performance degradation, confirming that these features retain essential predictive information. The model with the highest accuracy is thus selected as the base learner. Importantly, the same optimal feature subset is identified from the high-temperature data set, demonstrating that feature consistency across temperatures is a foundational premise for transfer learning. Given this consistency, the TL strategy is implemented as an incremental retraining process. Specifically, the base learner is initially trained on the RT source domain using the optimal feature subset. Subsequently, HT data are progressively incorporated into the training set in increasing proportions, and the model is retrained on the combined dataset at each step. This incremental integration strategy enables the model to gradually adapt to the target domain while retaining knowledge learned from the source domain.

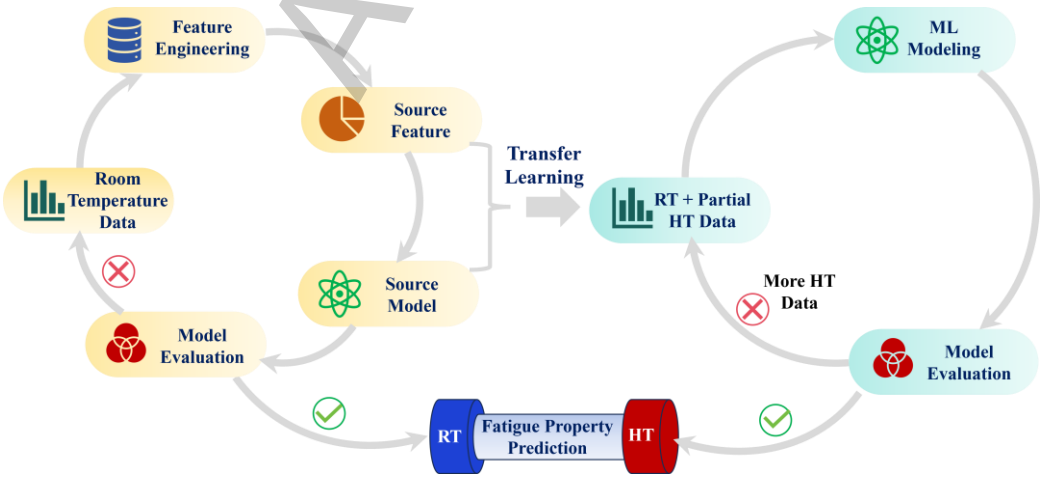

Fig. 1. A transfer learning framework for predicting high-temperature fatigue property using room-temperature fatigue data.

Data collection and processing

The experimental data used in this research are obtained from the MatNavi database, a publicly available collection of fatigue data sheets published by the National Institute of Materials Science [29]. Table 1 presents data on high-temperature and high-cycle fatigue performance for the steel grades under the specified fatigue testing conditions.

Table 1. Steel grades and corresponding fatigue testing conditions employed in this research

Steel grade	SUS403-B, S45C, SUS316-HP, SCM435, NCF800H, SUH616-B, SUS304-HP, SCMV4, ASTM A470-8
Type of fatigue test	rotating bending
Loading condition	load control under zero mean stress
Frequency	125 Hz
Stress concentration factor	1.0

Given the wide dispersion of fatigue life values relative to other parameters, the data are transformed using the base-10 logarithm for analysis. After data processing, two datasets representing fatigue performance at room-temperature (RT) and high-temperature (HT) are formed. The RT dataset consists of 60 data points, while the HT datasets at 400 ℃, 500 ℃, and 600 ℃ comprise 50, 59, and 45 data points, respectively. Each data point consists of 29 features, including processing details, chemical composition, mechanical properties, stress amplitude (σ), and target property (logNf). The chemical composition includes C, Si, Mn, P, S, Ni, Cr, Mo, W, etc., the processing details include ingot size (Is), reduction ratio (Rr), non-metallic inclusions (dA, dB, dC), austenite grain size number (Ags) and the mechanical properties include tensile strength (TS), 0.2% proof stress (YS), true fracture stress (TFS), elongation (EL) and reduction of area (RA). The distributions of these variables and the target attribute, fatigue life, are illustrated in Fig. 2. A summary of the feature variables, including their full descriptions and abbreviations, is presented in Table S1.

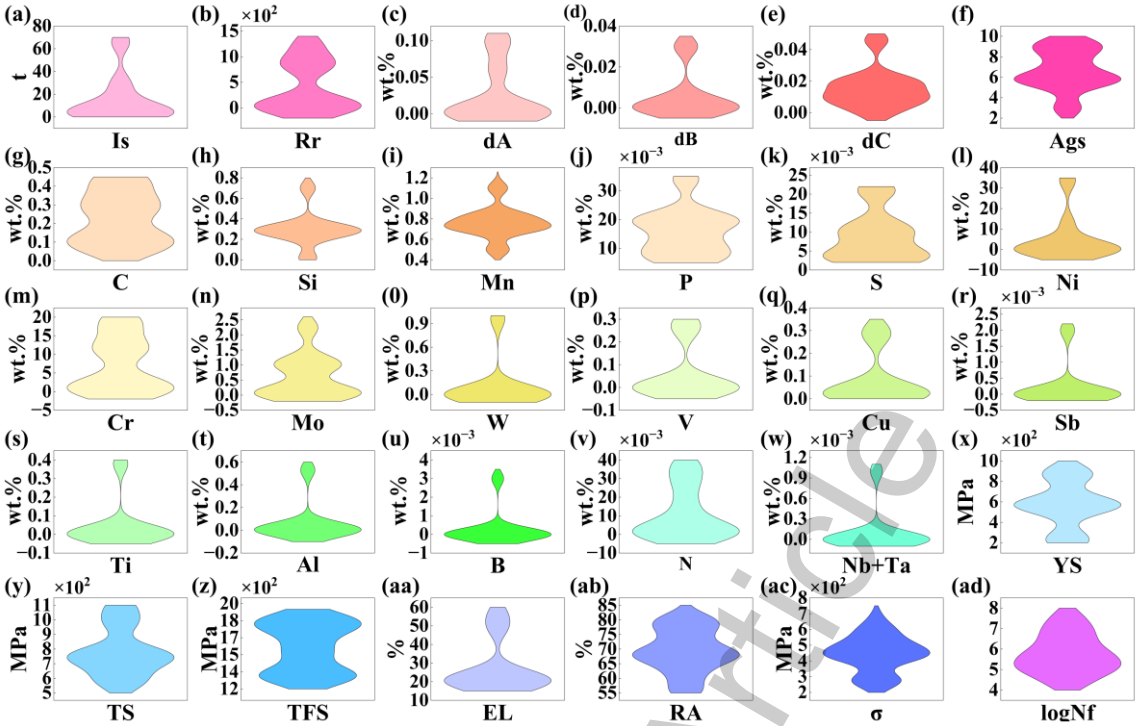

Fig. 2. Overview of feature variable and target performance distributions in the dataset.

Before performing ML, the collected data are pre-processed by normalizing each variable to the [0,1] interval using the following equation, which eliminates the influence of differing measurement scales on model predictive performance [30].

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

where x is the original data value, x' is the normalized result, $min(x)$ and $max(x)$ are the minimum and maximum values in the dataset, respectively.

Machine learning algorithms

In this study, six regression algorithms, including Ridge, Lasso, Random Forest (RF), Extreme Gradient Boosting (XGBoost), Categorical Boosting (CatBoost), and Gradient Boosting Decision Tree (GBDT), are used to model fatigue life. The six models are selected to provide a balanced comparison between interpretable linear methods and powerful nonlinear ensemble techniques. Ridge and Lasso are linear regression models with L2 and L1 regularization, respectively, that help prevent overfitting when dealing with high-dimensional features. However, their linear nature limits their ability to capture the complex nonlinear relationships

common in fatigue behavior. In contrast, RF and boosting algorithms, including XGBoost, CatBoost, and GBDT, are ensemble methods that can model nonlinear interactions without requiring explicit functional forms. RF reduces overfitting by averaging predictions from multiple decision trees. XGBoost efficiently handles missing values and uses regularization to control model complexity. CatBoost is specifically designed to handle categorical features and reduce overfitting. GBDT provides a flexible framework with robust performance across various datasets. Despite their strengths, ensemble methods are generally more computationally expensive and sensitive to hyperparameter tuning [31,32].

The efficacy of these models is assessed using four key evaluation metrics, namely the coefficient of determination (R^2) [33], Root Mean Square Error (*RMSE*) [34], Mean Absolute Error (*MAE*) [35], and Pearson correlation coefficient (*PCC*) [36]. R^2 quantifies the proportion of variance explained and serves as the primary metric for assessing a regression model's goodness of fit, with values closer to 1 indicating greater model accuracy. *MAE* measures the average magnitude of errors without emphasizing large deviations, offering robustness to outliers, while *RMSE* calculates the square root of the mean squared error and is more sensitive to large errors. For both metrics, smaller values indicate higher predictive accuracy [37,38]. In regression problems, the *PCC* can be used to evaluate a model's predictive accuracy by quantifying the linear correlation between the predicted and true values. It ranges from -1 to 1, with values close to 1 or -1 indicating a strong linear relationship and values near 0 indicating no linear correlation. These evaluation metrics are defined in Eqs. (2)-(5).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\frac{1}{n}(y_i - \bar{y})^2} \quad (2)$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{1}{n}(\hat{y}_i - y_i)^2} \quad (3)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (4)$$

$$PCC = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \quad (5)$$

where y_i, $\hat{y}_i$, $\bar{y}$, and $\bar{\hat{y}}$ denote the actual value, predicted value, the mean of actual values, and the mean of predicted values, respectively, with n representing the total number of samples.

To prevent overfitting while maximizing data utility, model evaluation is performed using 10-fold cross-validation [39]. This method divides the data into 10 subsets, each serving as the validation set in turn, with the remaining subsets forming the training set. This strategy is particularly beneficial for limited data, as it ensures that all data points participate in both training and validation phases, thereby enabling a thorough and reliable assessment of model performance. Each model underwent 50 independent training runs with diverse random seeds to enhance the stability and reliability of the results. The averaged R^2, *RMSE*, *MAE* and *PCC* values are reported, with the standard deviation across repeated runs used as error bars. The hyperparameters of each model are automatically optimized using Optuna, and the corresponding configurations for all models are listed in Table S2.

RESULTS AND DISCUSSION

Evaluation of RT models using the full dataset

The RT fatigue life is modeled using six regression algorithms with the full feature set. The evaluation metrics of these models are summarized in the radar chart shown in Fig. 3(a). The GBDT model demonstrates superior performance compared with the others. The scatter plot of experimental values versus average predicted values, along with the corresponding residual plot for the GBDT model, both with error bars, is presented in Fig. 3(b). The GBDT model achieves the highest R^2 of 0.895 (± 0.022) and *PCC* of 0.949 (± 0.011), and the lowest *RMSE* of 0.313 (± 0.030) and *MAE* of 0.230 (± 0.020). The model's high accuracy in predicting fatigue life in the RT dataset is evidenced by the close alignment of most data points with the line of best fit. The residual plot shows a relatively random distribution around the zero line, with residuals predominantly confined to ± 0.5 units across the experimental range. A slight heteroscedasticity is evident, with marginally greater dispersion at higher experimental logNf values (> 6.5), indicating increased prediction uncertainty in the upper regime. The near-symmetric scatter above and below the zero reference line corroborates the absence of significant prediction bias. The performance of the remaining models is shown in Fig. S1.

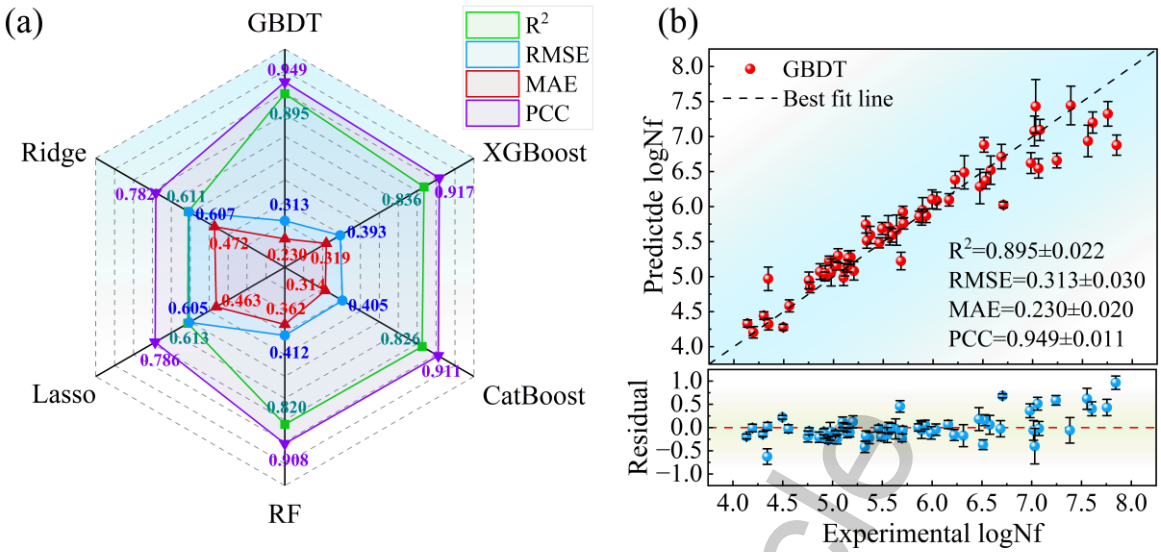

Fig. 3. Performance comparison of ML models using the complete dataset, with error bars representing the standard deviation. (a) Radar chart for the performance comparison. (b) The relationship between the actual values and the predicted values for the GBDT model, along with the corresponding residual plot. (including the mean results from 50 runs with different random seeds).

Feature engineering

Feature selection is a critical step in optimizing ML models, as it reduces complexity and computational cost by identifying and retaining only the features with the highest information content. This process not only improves model performance by mitigating overfitting but also enhances interpretability by clarifying the key factors driving predictions. Using SHAP (SHapley Additive exPlanations) [40] to rank and visualize feature importance significantly enhances the interpretability of fatigue life models. This method assigns each feature a consistent importance value based on its contribution to predictions, thereby uncovering key interactions between input variables and the target. The significance of these characteristics regarding fatigue life at RT is shown in Fig. 4, and the feature importance under HT is shown in Fig. S2.

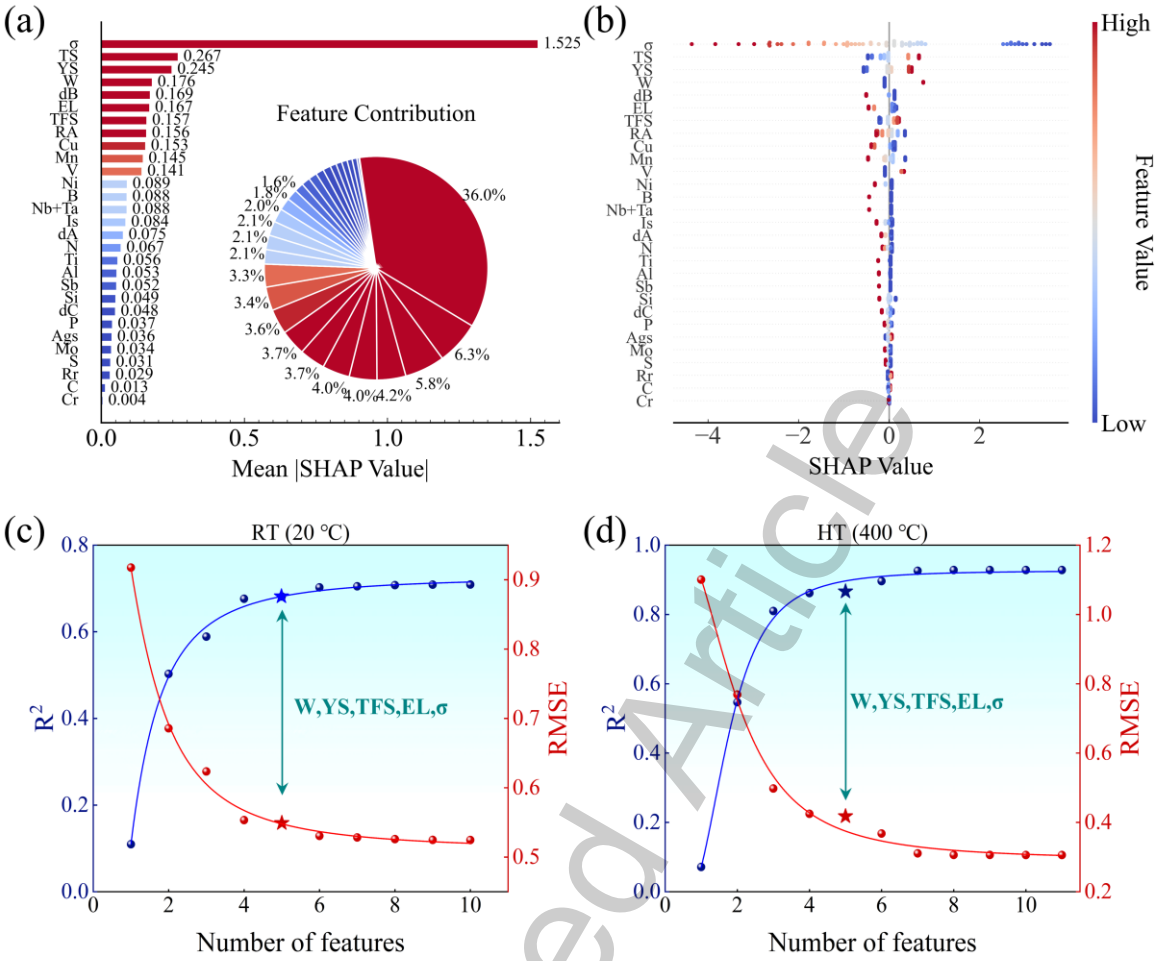

Fig. 4. Further feature selection for the ML models. (a) Bar chart of feature importance and ring chart of feature contribution ratios. (b) Summary plot of SHAP values. The optimal subset selection for the (c) room-temperature and (d) high-temperature dataset.

In Fig. 4(a–b), each point represents a SHAP value, and the color encodes the feature value according to the color bar (warmer colors indicate larger feature values, cooler colors indicate smaller feature values). When points with warmer colors cluster on the right (positive SHAP values), larger feature values tend to increase the prediction. The full ranking of feature importance guides the selection of the optimal subset. Feature importance analysis for fatigue life prediction identified σ, TS, YS, W, dB, EL, TFS, RA, Cu, and Mn as the ten most influential features, which were therefore selected for model inclusion. To further refine the feature set, the optimal subset method is applied [41]. This statistical approach evaluates all possible combinations of features to identify the subset that provides the best fit according to the specified model criteria. This method fits a least-squares model to every possible feature combination, from single features to the full feature set, and selects the best subset based on a

comprehensive performance comparison.

The result of the optimal subset selection for the RT dataset is shown in Fig. 4(c). Due to the large number of feature subsets, only the best subsets for each subset size are shown. The analysis indicates that model performance improves with an increasing number of features, as evidenced by a rise in R^2 and a corresponding decrease in *RMSE*. Both metrics stabilize after five features are included, suggesting that additional features contribute little to further improvement. An optimal feature subset consisting of five variables (W, YS, TFS, EL, and σ) is identified to balance predictive accuracy and model complexity. Through an exhaustive search, this subset is the optimal five-feature combination in terms of R^2 for the current dataset, and its robustness is confirmed by 10-fold cross-validation with 50 diverse random seeds. From a physical perspective, σ directly governs the applied stress level and is inversely correlated with fatigue life. YS and TFS reflect the material's resistance to plastic deformation and fracture, respectively, both of which are critical under cyclic loading. EL represents ductility, which influences fatigue crack initiation and propagation. The W content can improve fatigue performance through solid solution strengthening and precipitation hardening induced by fine particles, thereby enhancing the alloy's fatigue resistance [42]. The inclusion of these five features thus captures the key factors governing fatigue behavior while maintaining model simplicity.

Post-selection evaluation of RT models

The five features identified through feature selection in the previous section are used as additional inputs to the ML model. The performance of the diverse models is shown in the radar chart in Fig. 5(a). Compared with the case using all features, no significant degradation in model performance is observed, indicating the reasonableness of the feature engineering process. Among the six evaluated machine learning models, the GBDT model again demonstrated the highest prediction accuracy, achieving the highest R^2 (0.893 ± 0.023) and *PCC* (0.948 ± 0.012), along with the lowest *RMSE* (0.317 ± 0.032) and *MAE* (0.234 ± 0.020). The scatter plot of the GBDT model showing the relationship between the experimental and predicted values, along with the corresponding residual plot, is shown in Fig. 5(b). The predictions are in close agreement with the experimental values, with minimal errors. Therefore, the GBDT models developed in this study demonstrate satisfactory predictive capability and stability for fatigue life prediction within the compositional scope of the original dataset, as evidenced by R^2 values exceeding 0.85 and error bars smaller than 0.03. The residuals show a largely symmetric

distribution around the zero baseline, with most data points clustered within a narrow ± 0.5 interval, indicating robust model generalization. A mild asymmetry emerges at the extremes of the dataset: reduced residual variance is observed in the lower range where experimental logNf falls below 4.5, whereas greater variability with occasional positive outliers appears in the upper range where experimental logNf exceeds 7.0, suggesting reduced predictive accuracy under extreme experimental conditions. Throughout the central operating window spanning experimental logNf from approximately 5.5 to 6.8, minimal directional bias is observed, and no discernible correlation structure is present among the residuals. Additionally, the detailed performance of the remaining models on the RT dataset after feature selection is presented in Fig. S3.

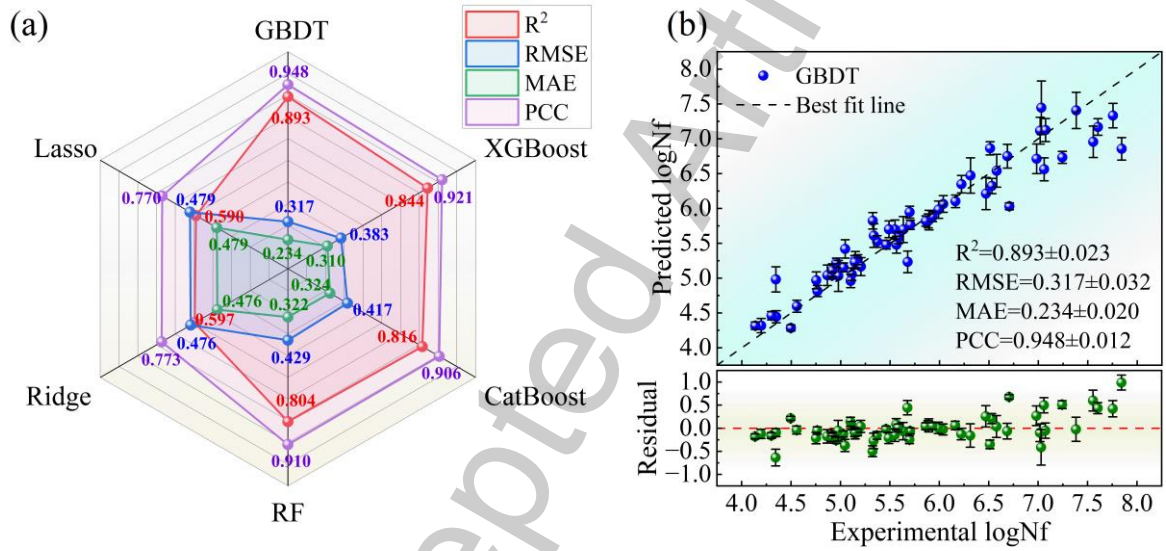

Fig. 5. Comparison of model performance after feature selection, with error bars representing the standard deviation. (a) Radar chart for the performance comparison. (b) The relationship between the actual values and the predicted values for the GBDT model, along with the corresponding residual plot. (including the mean results from 50 runs with different random seeds).

HT models transfer prediction

Directly applying the GBDT model trained on RT data to predict HT fatigue life yields poor results and shows virtually no predictive capability. This discrepancy underscores the model's inadequacy in predicting HT behavior when trained solely on RT data. Physically, HT exposure induces substantial microstructural evolution in steels, including recovery, recrystallization, precipitate coarsening, and dislocation structure rearrangement. Consequently, fatigue

resistance and failure mechanisms differ fundamentally between RT and HT conditions. Even when the chemical composition and loading conditions remain identical, the underlying physical processes governing fatigue life are not directly transferable. Statistically, the relationship between input features and fatigue life shows a domain shift from RT to HT, indicating that the feature-target distributions in the two domains are misaligned. To bridge this domain gap, the TL strategy detailed in Section 2 is employed. In this framework, RT fatigue life serves as the source domain, while HT fatigue life constitutes the target domain. To implement TL effectively, feature selection is first performed on the HT data to align feature importance with that of the RT dataset. Using the optimal subset method, the key features for HT fatigue life are identified. As illustrated in Fig. 4(d), the optimal feature set comprising W, YS, TFS, EL, and σ aligns closely with the feature set derived from RT data. Furthermore, feature selection was performed at elevated temperatures of 500 ℃ and 600 ℃, and the results exhibited the same characteristics as those observed in the RT dataset, as shown in Fig. S4. Building on these selected features, TL is applied by training a model on RT fatigue data to predict HT fatigue life, thereby mitigating the challenge of limited HT data availability. By incrementally increasing the proportion of HT data during training, the model is continuously refined to better capture material behavior under elevated temperatures, which reduces prediction discrepancies and improves generalization across varying thermal conditions.

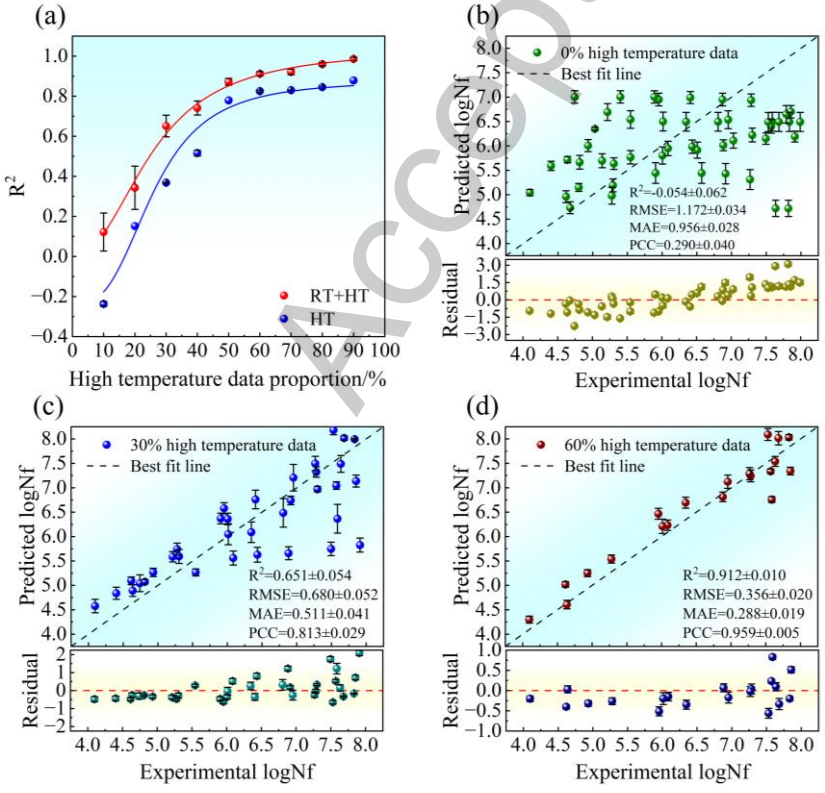

Fig. 6. The influence of high-temperature data proportion on the fatigue life prediction precision of the GBDT model, with error bars representing the standard deviation. (a) Comparison of R^2 values for GBDT models trained on mixed versus high-temperature only data. (b-d) The performance of the remaining high-temperature test set, containing 0%, 30%, and 60% of the high-temperature data.

The performance of the GBDT model is systematically evaluated by progressively incorporating HT data into the RT training set in 10% increments, from 10% to 90%. The remaining HT data are reserved as a test set. Additionally, the efficacy of the TL strategy is verified using a dataset containing only HT data for comparison. To ensure robustness, each configuration is trained 50 times using different random seeds. As illustrated in Fig. 6(a), the average R^2 values from both evaluation strategies increase monotonically as HT data are incorporated, indicating that the model progressively adapts to the target domain. Specifically, the R^2 value is 0.368 for the model trained on 30% HT data and 0.825 for the model trained on 60% HT data. More importantly, across all HT data fractions, the model trained on the combined RT-HT dataset consistently outperforms the model trained solely on HT data, demonstrating that knowledge learned from the data-rich RT domain provides a beneficial initialization and significantly enhances predictive capability under limited HT data conditions. This result highlights the necessity of the TL strategy, particularly when HT data are scarce. The GBDT model achieves the best performance on the RT dataset and is therefore initially considered suitable for TL from RT to HT. However, this remains an assumption, so TL from RT to HT is performed for all models as shown in Fig. S5-S9, and the GBDT model is ultimately selected after a comprehensive comparison of their performance.

The averaged performance over 50 iterations for mixing ratios of 0% 30% and 60% high-temperature data is shown in Fig. 6(b-d), respectively. When no HT data are included (Fig. 6b), the model exhibits poor predictive capability, with negative R^2 values, indicating a difference between the RT and HT domains. Incorporating 30% HT data yields moderate predictive accuracy (R^2 = 0.651), while increasing the fraction to 60% significantly improves performance (R^2 = 0.912), accompanied by substantial reductions in *RMSE* and *MAE* and a marked increase in *PCC*. This improvement is further corroborated by the evolution of the residual distributions. As shown in Fig. 6(b–d), the residuals progressively transition from systematic bias to stochastic dispersion as HT data increase. At 0% HT data, residuals show pronounced negative deviations across the entire prediction range, indicating a systematic overestimation of fatigue

life. With 30% HT data, the residual distribution becomes approximately symmetric around zero, although noticeable variance and heteroscedasticity persist, particularly at higher fatigue life levels. When the HT fraction reaches 60%, residuals collapse into a narrow and homoscedastic band, indicating stable and unbiased predictions. The progressive performance improvement with increasing HT data reflects the model's enhanced adaptation to the target domain through the incorporation of HT information. Transfer predictions are further investigated across temperatures ranging from RT to 500 °C and from RT to 600 °C, as presented in Fig. S10 and Fig. S11, respectively. These results not only confirm the effectiveness of the TL strategy but also provide empirical evidence for consistent fatigue behavior patterns across temperatures, thereby fundamentally justifying the applicability of TL in this context.

The present framework is fundamentally data-driven and does not explicitly incorporate microstructural evolution or temperature-dependent deformation mechanisms. Although consistent feature subsets are identified across RT and HT datasets, this consistency reflects similarity in predictive relevance rather than mechanistic equivalence. The failure of direct RT-to-HT prediction indicates a significant domain gap, whereas improved performance with increasing HT data suggests that model adaptation relies on partial retraining. Therefore, the proposed approach should be understood as a transfer-assisted learning strategy that improves data efficiency rather than a strict transfer of physical knowledge. HT fatigue involves additional mechanisms such as creep-fatigue interaction, oxidation-assisted damage, and microstructural instability, which are not explicitly captured in the present framework due to data limitations [43]. Instead, their effects are indirectly embedded in macroscopic descriptors such as strength and elongation.

The input features include compositional, processing, and macroscopic mechanical descriptors, which are used to establish predictive relationships with fatigue life. These descriptors are treated as static, global variables and do not explicitly capture microstructural evolution or local damage processes during fatigue. Although microstructural features such as grain size and inclusion characteristics are included, their roles are represented only through statistical correlations rather than explicit mechanistic modeling. Consequently, SHAP-based feature importance reflects predictive relevance rather than causal relationships between microstructure, deformation, and fatigue damage evolution. The observed feature consistency across temperatures therefore indicates similarity in statistical patterns within the dataset, rather than invariance of the underlying physical mechanisms. Given the current data availability,

macroscopic properties such as strength and elongation can be regarded as integrated surrogate descriptors that implicitly capture the combined influence of microstructure and its evolution. However, incorporating microstructure-sensitive descriptors and explicitly modeling temperature-dependent microstructure evolution remain necessary to improve physical interpretability and extrapolation capability in future work. Importantly, the proposed TL framework is developed and validated specifically for alloy steels under the testing conditions summarized in Table 1 (rotating bending, load control with zero mean stress, 125 Hz).

CONCLUSION

In this study, comprehensive datasets describing the fatigue properties of steels at both RT and HT are compiled. ML model development and feature selection are first conducted using the RT dataset to identify critical variables governing fatigue life. Subsequently, a TL strategy is implemented to enable HT fatigue life prediction by leveraging knowledge learned from RT data and incorporating a limited amount of HT data for incremental retraining. This method utilized knowledge from RT data by incorporating a limited portion of HT data into the training process, thereby significantly enhancing the prediction accuracy of the GBDT model for HT fatigue life. Using this applied methodology, the method validates the effectiveness of predicting material fatigue properties across different temperatures and confirms the consistency of material properties across various temperature ranges via the optimal subset method. The study establishes a practical framework for material selection and design in HT applications and paves the way for future predictive modeling of material performance under a variety of operating conditions.

DECLARATIONS

Authors' contributions

Writing - original draft, software, methodology, formal analysis, data curation: Qi Wang, Chunlei Shang

Writing - review and editing, supervision, project administration, investigation: Hong-Hui Wu, Dexin Zhu

Validation, investigation, data collection: Shuize Wang, Junheng Gao

Visualization, resources, formal analysis, conceptualization: Haitao Zhao, Chaolei Zhang, Yuhe Huang

Supervision, conceptualization, revised and finalized the manuscript: Ju Lu, Xinping Mao

Availability of data and materials

The dataset used in this study is obtained from the NIMS Structural Materials Data Sheet Service (available at https://fds.nims.go.jp/), specifically focusing on the "Elevated-Temperature, High-Cycle Fatigue Properties" data sheets. The processed data that support the findings of this study are available from the corresponding author upon reasonable request. Additionally, the source code and project files used in this research have been made available in the following GitHub repository: https://github.com/wq360/project.git.

AI and AI-assisted tools statemen

Not applicable.

Financial support and sponsorship

This work is financially supported by the Advanced Materials-National Science and Technology Major Project: 2025ZD0619601. Hong-Hui Wu also thanks the financial support from the Xiaomi Young Scholars Program, and the Xiaomi Open-Competition Research Program (39990320). The computing work is supported by USTB MatCom of Beijing Advanced Innovation Center for Materials Genome Engineering.

Conflicts of interest

Hong-Hui Wu is a Youth Editorial Board Member of the journal *Journal of Materials Informatics*, and a Guest Editor of the Special Topic "Predictive Materials Informatics for Extreme Conditions", but was not involved in any steps of editorial processing, notably including reviewer selection, manuscript handling, and decision making, while the other authors have declared that they have no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Li, Y.; Liu, J.X.; Huang, W.Q.; Zhang, S. Microstructure related analysis of tensile and fatigue properties for sand casting aluminum alloy cylinder head. *Eng. Fail. Anal.* **2022**, *136*, 106210. DOI: https://doi.org/10.1016/j.engfailanal.2022.106210

2. Schütz, W. A history of fatigue. *Eng. Fract. Mech.* **1996**, *54*, 263–300. DOI: https://doi.org/10.1016/0013-7944(95)00178-6

3. Peng, W.J.; Xue, H.; Ge, R.; Peng, Z. The influential factors on very high cycle fatigue testing results. In Proceedings of the 12th International Fatigue Congress (FATIGUE), Poitiers, FRANCE, 2018; pp. 9–11. DOI: https://doi.org/10.1051/matecconf/201816520002

4. Yang, G.X.; Wang, M.; Li, Q.; Ding, R. Methodology to Evaluate Fatigue Damage of High-Speed Train Welded Bogie Frames Based on On-Track Dynamic Stress Test Data. *Chin. J. Mech. Eng.* **2019**, *32*, 51. DOI: https://doi.org/10.1186/s10033-019-0365-3

5. Jimenez-Martinez, M. Manufacturing effects on fatigue strength. *Eng. Fail. Anal.* **2019**, *108*, 104339. DOI: https://doi.org/10.1016/j.engfailanal.2019.104339

6. Dixon, W.J.; Mood, A.M. A Method for Obtaining and Analyzing Sensitivity Data. *J. Am. Stat. Assoc.* **1948**, *43(241)*, 109–126. DOI: https://doi.org/10.1080/01621459.1948.10483254

7. Veile, G.; Rudolph, J.; Grozinger, N.; Herzig, M.; Grimm, M.; Weihe, S. Improving fatigue testing of AISI 304L stainless steel in high temperature water regarding their complex hardening and softening material behaviou. *Int. J. Pressure. Vessels. Pip*. **2025**, *218*, 105612. DOI: https://doi.org/10.1016/j.ijpvp.2025.105612

8. Hirose, T.; Sakasegawa, H.; Nakajima, M.; Kato, T.; Miyazawa, T.; Tanigawa, H. Effects of test environment on high temperature fatigue properties of reduced activation ferritic/martensitic steel, F82H. *Fusion Eng. Des.* **2018**, *136*, 1073–1076, DOI: https://doi.org/10.1016/j.fusengdes.2018.04.072

9. Kamal, M.; Rahman, M. Advances in fatigue life modeling: A review. *Renew. Sust. Energ. Rev.* **2018**, *82*, 940–949. DOI: https://doi.org/10.1016/j.rser.2017.09.047

10. Zhu, S.P.; Yu, Z.Y.; Correia, J.; De Jesus, A.; Berto, F. Evaluation and comparison of critical plane criteria for multiaxial fatigue analysis of ductile and brittle materials. *Int. J. Fatigue.* **2018**, *112*, 279–288. DOI: https://doi.org/10.1016/j.ijfatigue.2018.03.028

11. Fatemi, A.; Socie, D.F. A critical plane approach to multiaxial fatigue damage including out-of-phase loading. *Fatigue. Fract. Engng. Mater. Struct*. **1988**, *11*, 149–165. DOI: https://doi.org/10.1111/J.1460-2695.1988.TB01169.X

12. Wang, H.J.; Liu, X.T.; Chen, T.; Xu, S. Prediction and evaluation of fatigue life via modified energy method considering surface processing. *Int. J. Damage Mech.* **2022**, *31*, 426–443. DOI: https://doi.org/10.1177/10567895211045120

13. Cui, W.C. A state-of-the-art review on fatigue life prediction methods for metal structures. *J. Mar. Sci. Technol.* **2002**, *7*, 43–56. DOI: https://doi.org/10.1007/s007730200012

14. Shang, C.L.; Jiang, T.B.; Wu, H.H.; et al. Efficient design of hydrogen-resistant ultra-high-strength steels via active learning and multiscale characterization. *Corros. Commun.* **2026**, *21*, 60-72. DOI: https://doi.org/10.1016/j.corcom.2025.12.002

15. Zhang, M.; Sun, C.N.; Zhang, X.; et al. High cycle fatigue life prediction of laser additive manufactured stainless steel: A machine learning approach. *Int. J. Fatigue* **2019**, *128*, 105194. DOI: https://doi.org/10.1016/j.ijfatigue.2019.105194

16. Yang, J.Y.; Kang, G.Z.; Liu, Y.J.; Kan, Q.H. A novel method of multiaxial fatigue life prediction based on deep learning. *Int. J. Fatigue.* **2021**, *151*, 106356. DOI: https://doi.org/10.1016/j.ijfatigue.2021.106356

17. Arvanitis, K.; Nikolakopoulos, P.; Pavlou, D.; Farmanbar, M. Machine learning-based fatigue lifetime prediction of structural steels. *Alex. Eng. J.* **2025**, *125*, 55–66. DOI: https://doi.org/10.1016/j.aej.2025.04.014

18. Freed, Y. Machine Learning-Based predictions of crack growth rates in an aeronautical aluminum alloy. *Theor. Appl. Fract. Mech.* **2024**, *130*, 104278. DOI: https://doi.org/10.1016/j.tafmec.2024.104278

19. Rathore, S.; Kumar, A.; Kumar, A.; Mishra, K.; Singh, A. Prediction of sub-critical fatigue crack growth rate in a high-carbon tempered martensitic steel at varying R ratio: Experimental investigation and machine learning based modelling. *Int. J. Fatigue.* **2025**, *193*, 108804. DOI: https://doi.org/10.1016/j.ijfatigue.2025.108804

20. Shiraiwa, T.; Miyazawa, Y.; Enoki, M. Prediction of Fatigue Strength in Steels by Linear Regression and Neural Network. *Mater. Trans.* **2019**, *60*, 189–198, DOI: https://doi.org/10.2320/matertrans.ME201714

21. Choi, D. Data-Driven Materials Modeling with XGBoost Algorithm and Statistical Inference Analysis for Prediction of Fatigue Strength of Steels. *Int. J. Precis. Eng. Manuf.* **2019**, *20*, 129–138. DOI: https://doi.org/10.1007/s12541-019-00048-6

22. Yan, F.; Song, K.; Liu, Y.; Chen, S.W.; Chen, J.Y. Predictions and mechanism analyses of the fatigue strength of steel based on machine learning. *J. Mater. Sci.* **2020**, *55*, 15334–15349. DOI: https://doi.org/10.1007/s10853-020-05091-7

23. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22(10)*, 1345–1359. DOI: https://doi.org/10.1109/TKDE.2009.191

24. Tang, S.N.; Ma, J.T.; Yan, Z.Q.; Zhu, Y.; Khoo, B. Deep transfer learning strategy in intelligent fault diagnosis of rotating machinery. *Eng. Appl. Artif. Intell.* **2024**, *134*, 108678. DOI: https://doi.org/10.1016/j.engappai.2024.108678

25. Hu, J.T.; Chen, M.; Tang, H.L.; Zhang, J.Y. An adversarial transfer learning method based on domain distribution prediction for aero-engine fault diagnosis. *Eng. Appl. Artif. Intell.* **2024**, *133*, 108287. DOI: https://doi.org/10.1016/j.engappai.2024.108287

26. Li, Y.; Wei, P.T.; Xiang, G.; Jia, C.F.; Liu, H.J. Gear contact fatigue life prediction based on transfer learning. *Int. J. Fatigue.* **2023**, *173*, 107686. DOI: https://doi.org/10.1016/j.ijfatigue.2023.107686

27. Wei, X.L; Zhang, C.; Han, S.Y.; Jia, Z.X.; Wang, C.C.; Xu, W. High cycle fatigue S-N curve prediction of steels based on transfer learning guided long short term memory network. *Int. J. Fatigue.* **2022**, *163*, 107050. DOI: https://doi.org/10.1016/j.ijfatigue.2022.107050

28. Zhai, Q.Q.; Liu, Z.; Zhu, P. A transfer learning enhanced physics-informed neural network for predicting fatigue life of megacasting alloy with limited data sizes. *Int. J. Fatigue.* **2025**, *200*, 109129. DOI: https://doi.org/10.1016/j.ijfatigue.2025.109129

29. Furuya, Y.; Nishikawa, H.; Hirukawa, H.; Nagashima, N.; Takeuchi, E. Catalogue of NIMS fatigue data sheets. *Science and Technology of Advanced Materials* **2019**, *20*, 1055–1072. DOI: https://doi.org/10.1080/14686996.2019.1680574

30. Patro, S.G.K.; sahu, K.K. Normalization: A Preprocessing Stage. *Iarjset* **2015**, 20–22. DOI: https://doi.org/10.48550/arXiv.1503.06462

31. Miao, B.; Lin, G.-Q.; Zhang, Y.; Cheng, Y.-W.; Yang, H. Machine learning applications in metallic materials: Recent advances and future perspectives. *J. Alloy. Compd.* **2026**, *1062*, 187486. DOI: https://doi.org/10.1016/j.jallcom.2026.187486

32. Shmuel, A.; Glickman, O.; Lazebnik, T. A comprehensive benchmark of machine and deep learning models on structured data for regression and classification. *Neurocomputing* **2025**, *655*, 131337. DOI: https://doi.org/10.1016/j.neucom.2025.131337

33. Colin Cameron, A.; Windmeijer, F.A.G. An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics* **1997**, *77*, 329–342. DOI: https://doi.org/10.1016/S0304-4076(96)01818-0

34. Herlocker, J.L.; Konstan, J.A.; Terveen, L.G.; Riedl, J.T. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* **2004**, *22*, 5–53. DOI: https://doi.org/10.1145/963770.963772

35. Cleger-Tamayo, S.; Fernández-Luna, J.M.; Huete, J.F. On the Use of Weighted Mean Absolute Error in Recommender Systems. In Proceedings of the RUE@RecSys, 2012.

36. Sheugh, L.; Alizadeh, S.H. A note on Pearson Correlation Coefficient as a metric of similarity in recommender system. *2015 Ai & Robotics (Iranopen)* **2015**. DOI: https://doi.org/10.1109/RIOS.2015.7270736

37. Chai, T.; Draxler, R. Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. DOI: https://doi.org/10.5194/gmd-7-1247-2014

38. Shang, C.L.; Zhu, D.Z.; Wu, H.H.; et al. A quantitative relation for the ductile-brittle transition temperature in pipeline steel. *Scr. Mater.* **2024**, *244*, 116023. DOI: https://doi.org/10.1016/j.scriptamat.2024.116023

39. Efron; B. Estimating the Error Rate of a Prediction Rule: Improvement on Cross-Validation. *Publications of the American Statistical Association* **1983**, *78*, 316–331. DOI: https://doi.org/10.1080/01621459.1983.10477973

40. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017; pp. 4768–4777. DOI: https://doi.org/10.48550/arXiv.1705.07874

41. Kohavi, R.; John, G.H. Wrappers for Feature Subset Selection. *Artif. Intell.* **1997**, *97*, 273–324. DOI: https://doi.org/10.1016/S0004-3702(97)00043-X

42. Park, J.S.; Kim, S.J.; Lee, C.S. Effect of W addition on the low cycle fatigue behavior of high Cr ferritic steels. *Mater. Sci. Eng. A-Struct. Mater. Prop. Microstruct. Process.* **2001**, *298*, 127–136. DOI: https://doi.org/10.1016/S0921-5093(00)01291-0

43. Lu, X.C.; Xu, Y.L.; Ran, H.; et al. The role of layer strength ratio in enhancing strain hardening and achieving strength-ductility synergy in heterostructured materials. *Acta Mater.* **2025**, *289*, 120928. DOI: https://doi.org/10.1016/j.actamat.2025.120928

Supplementary Materials

Table S1. Definitions of abbreviations and variables used in the study

	Abbreviation	Description
Inputs	Is	Ingot size (t)
	Rr	Reduction ratio (%)
	dA	For inclusions deformed by plastic work (%)
	dB	For inclusions occuring in discontinuous (%)
	dC	For isolated inclusions (%)
	Ags	Austenite grain size number
	C	Mass fraction of C (%)
	Si	Mass fraction of Si (%)
	Mn	Mass fraction of Mn (%)
	P	Mass fraction of P (%)
	S	Mass fraction of S (%)
	Ni	Mass fraction of Ni (%)
	Cr	Mass fraction of Cr (%)
	Mo	Mass fraction of Mo (%)
	W	Mass fraction of W (%)
	V	Mass fraction of V (%)
	Cu	Mass fraction of Cu (%)
	Sb	Mass fraction of Sb (%)
	Ti	Mass fraction of Ti (%)
	Al	Mass fraction of Al (%)
	B	Mass fraction of B (%)
	N	Mass fraction of N (%)
	Nb+Ta	Mass fraction of Nb+Ta (%)
	YS	0.2% proof stress (MPa)
	TS	Tensile strength (MPa)
	TFS	True fracture stress (MPa)
	EL	Elongation (%)
	RA	Reduction of area (%)
	σ	Stress amplitude (MPa)
Output	logNf	Number of cycles to failure

Table S2. Hyperparameters of the evaluated models

Models	Hyperparameters
GBDT	n estimators: 500 learning rate: 0.066 max depth: 10 min samples split: 4 min samples leaf: 1 subsample: 0.59
XGBoost	n estimators: 400 learning rate: 0.049 max depth: 10 subsample: 0.515 colsample bytree: 0.882 reg alpha: 0.0023 reg lambda: 0.161 min child weight: 1 gamma: 0.0053
CatBoost	n estimators: 400, learning rate: 0.0874 depth: 5 l2_leaf_reg: 0.0065 border count: 222 random strength: 0.155 bagging temperature: 0.0643
RF	n estimators: 350 max depth: 10 min samples split: 2 min samples leaf: 1
Lasso	alpha: 0.01 max iter: 1000
Ridge	alpha: 0.5 tol: 0.001

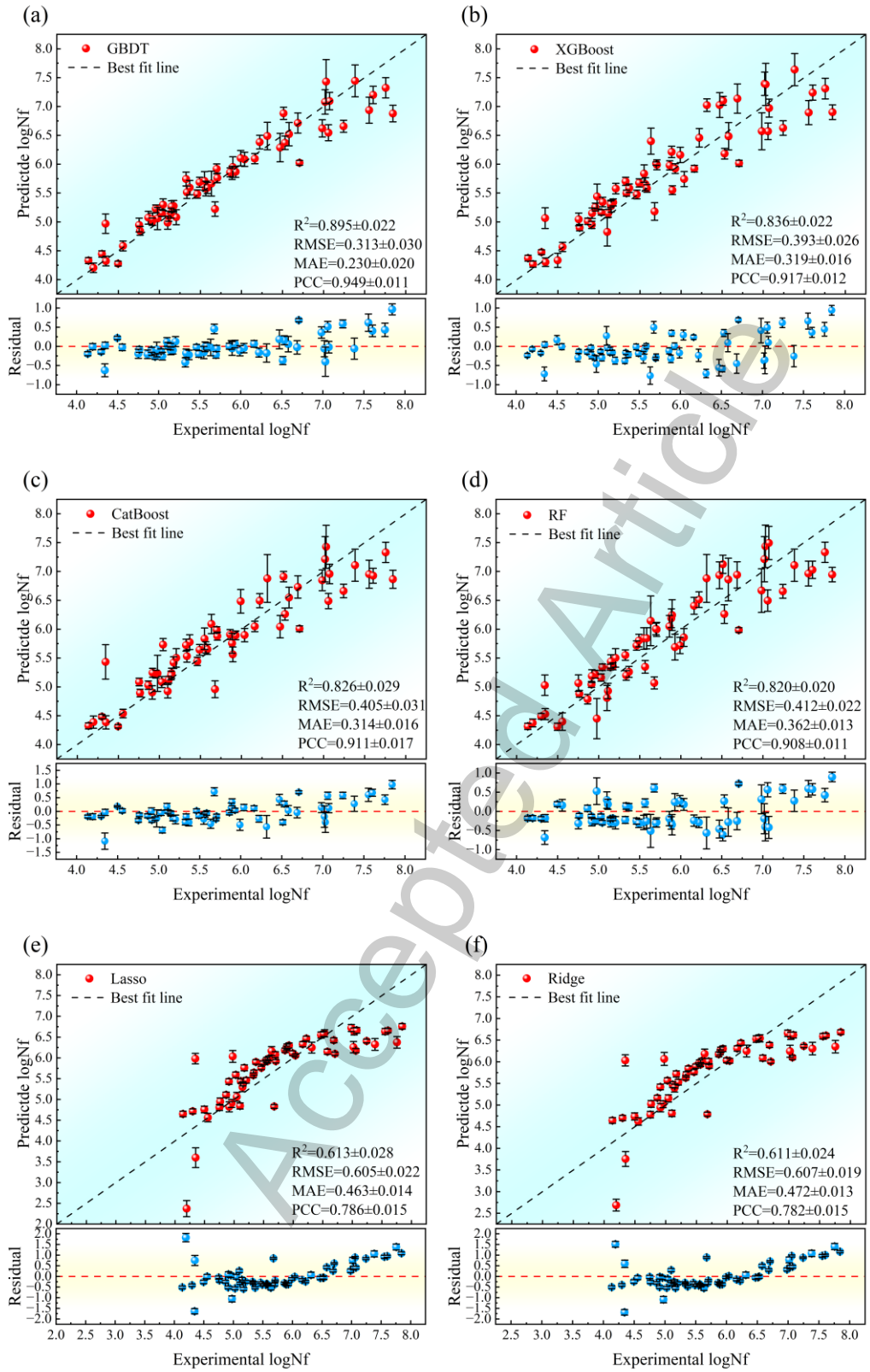

Fig. S1 The performance of the ML models on the full RT dataset, including scatter plots of actual versus predicted values and residual plots. (a) GBDT, (b) XGBoost, (c) CatBoost, (d) RF, (e) Lasso, and (f) Ridge.

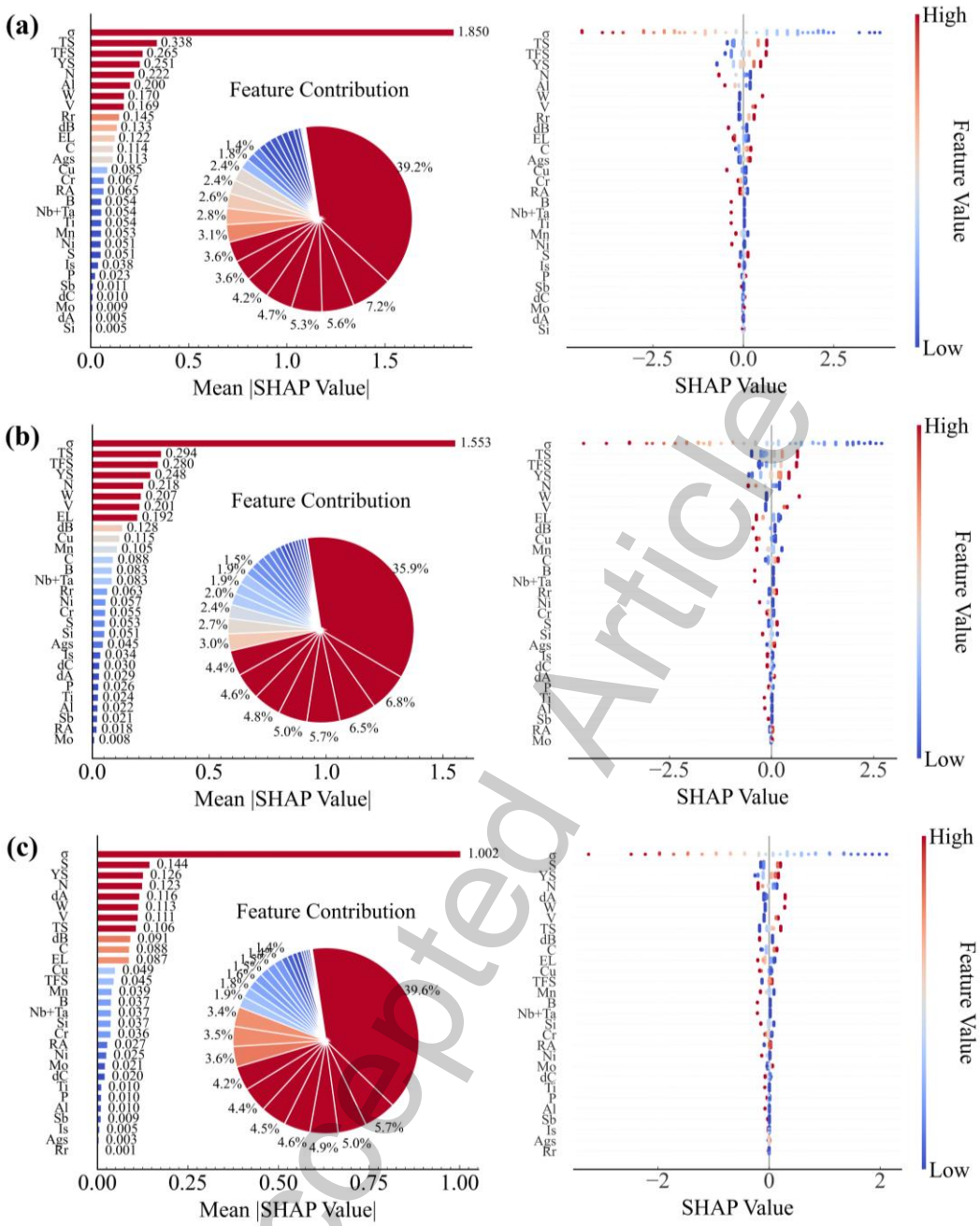

Fig. S2. Feature importance under high-temperature dataset. (a-c) correspond to the rankings of feature importance based on SHAP values at 400 °C, 500 °C, and 600 °C, respectively.

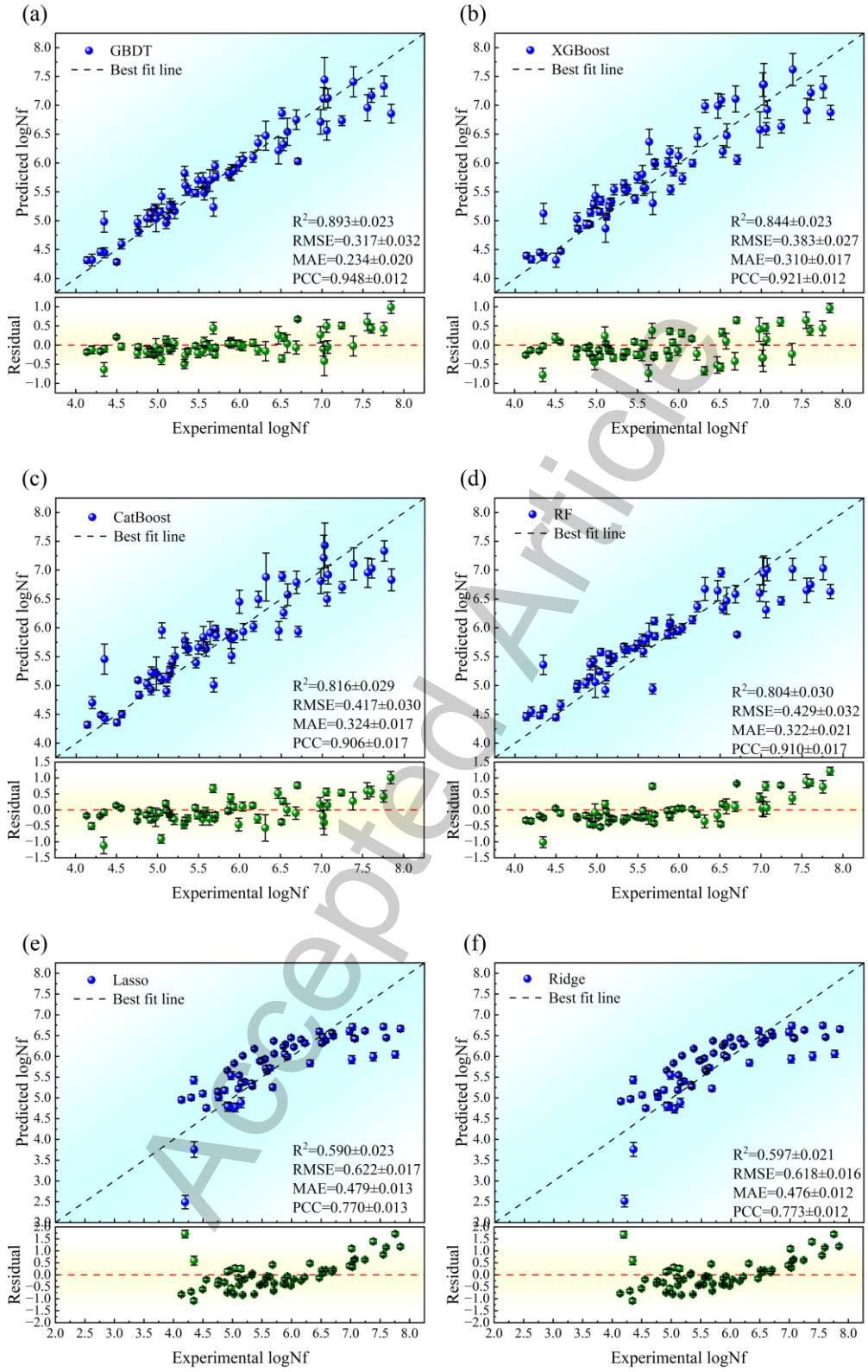

Fig. S3. The performance of the ML models on the RT dataset after feature selection, including scatter plots of actual versus predicted values and residual plots. (a) GBDT, (b) XGBoost, (c) CatBoost, (d) RF, (e) Lasso, and (f) Ridge.

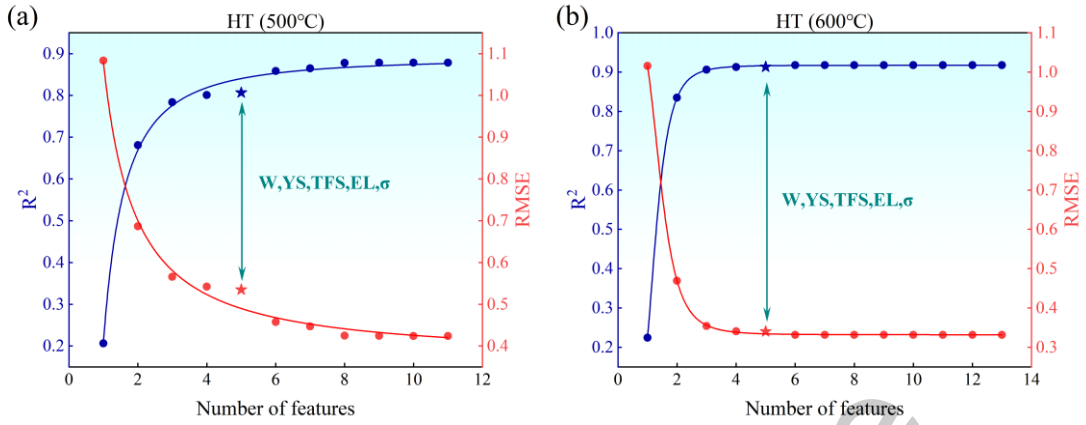

Fig. S4. Feature selection under high-temperature dataset. (a-b) represent the optimal subset selection at 500 ℃ and 600 ℃, respectively.

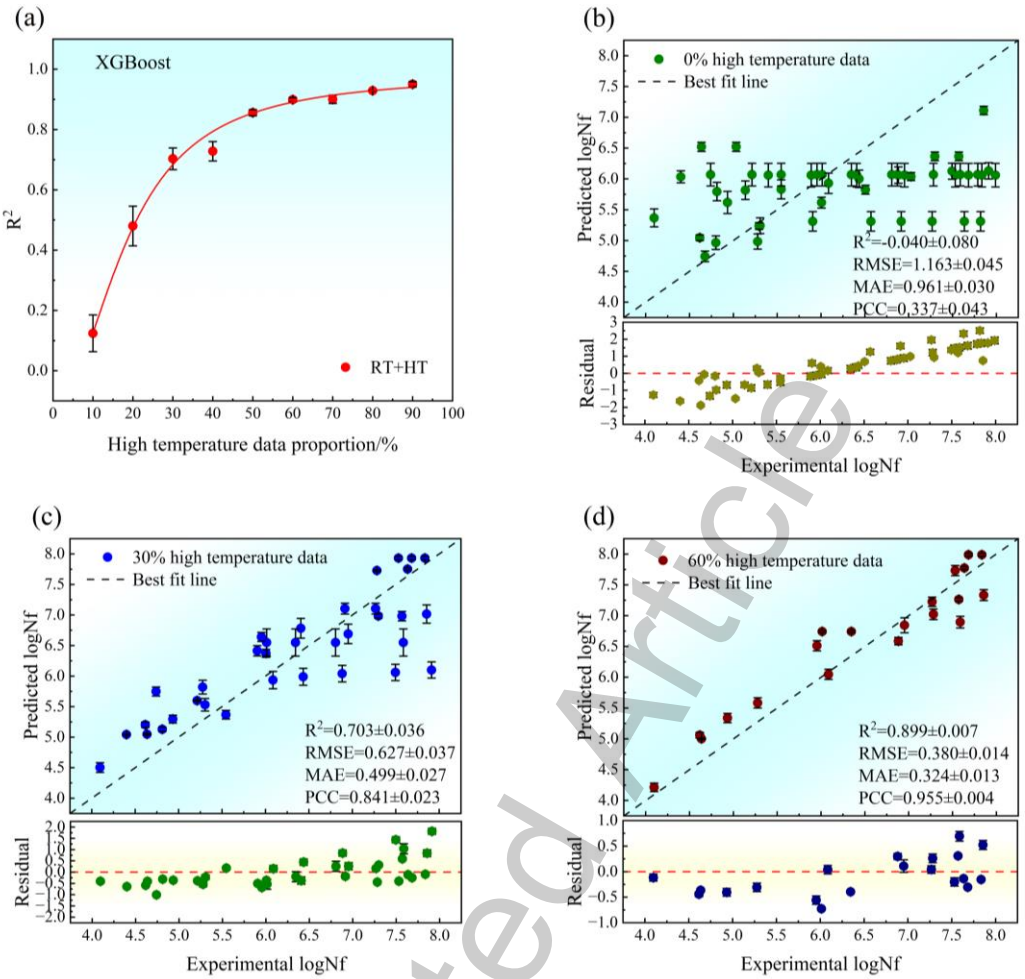

Fig. S5. Transfer learning prediction results of the XGBoost model from room temperature to high temperature. (a) R^2 values of the model under different proportions of high-temperature data. (b–d) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0 %, 30 %, and 60 %, respectively.

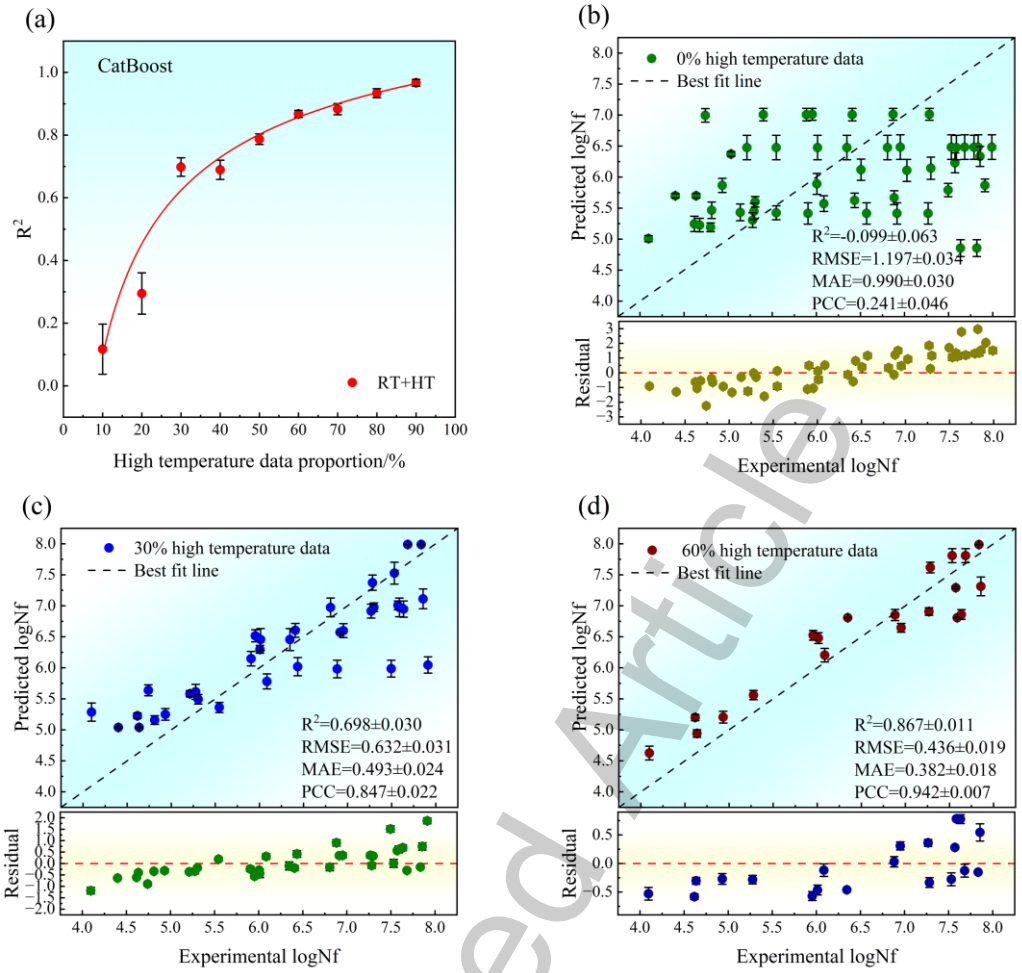

Fig. S6. Transfer learning prediction results of the CatBoost model from room temperature to high temperature. (a) R^2 values of the model under different proportions of high-temperature data. (b–d) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0 %, 30 %, and 60 %, respectively.

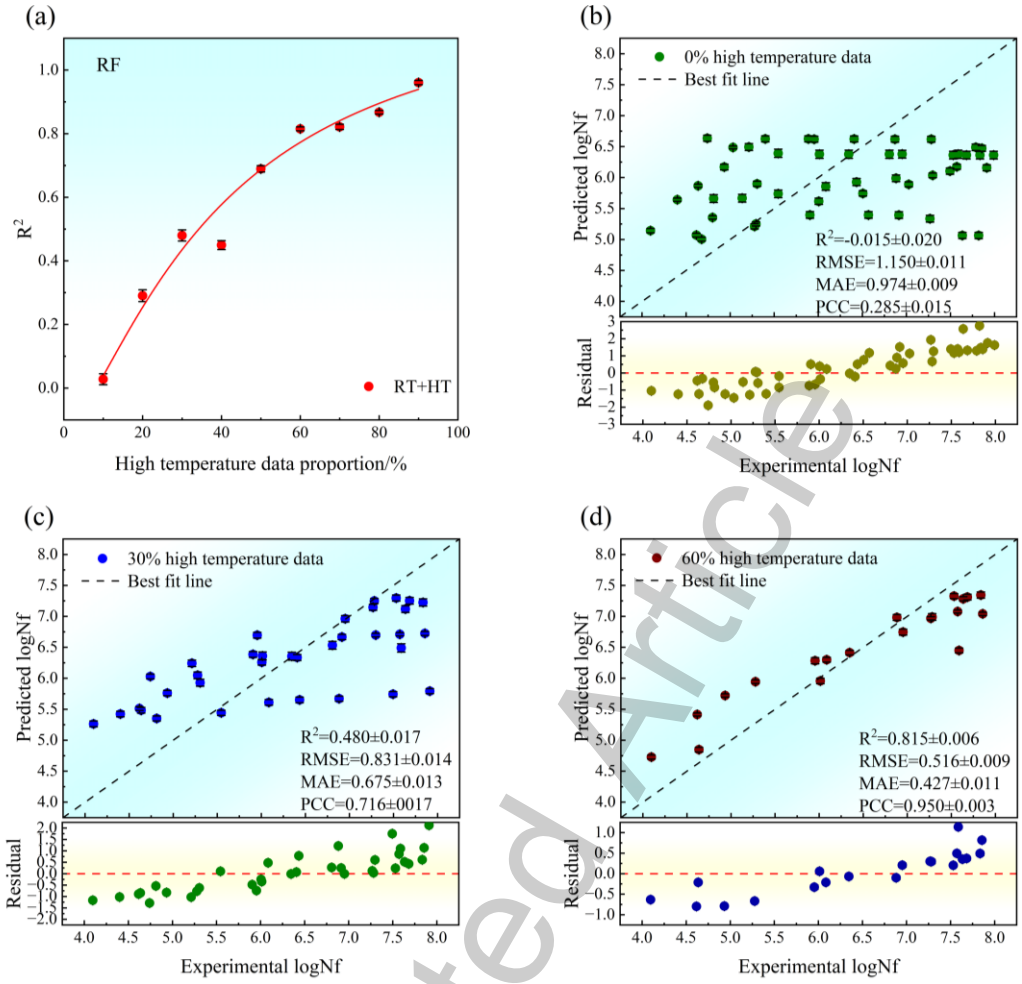

Fig. S7. Transfer learning prediction results of the RF model from room temperature to high temperature. (a) R^2 values of the model under different proportions of high-temperature data. (b–d) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0 %, 30 %, and 60 %, respectively.

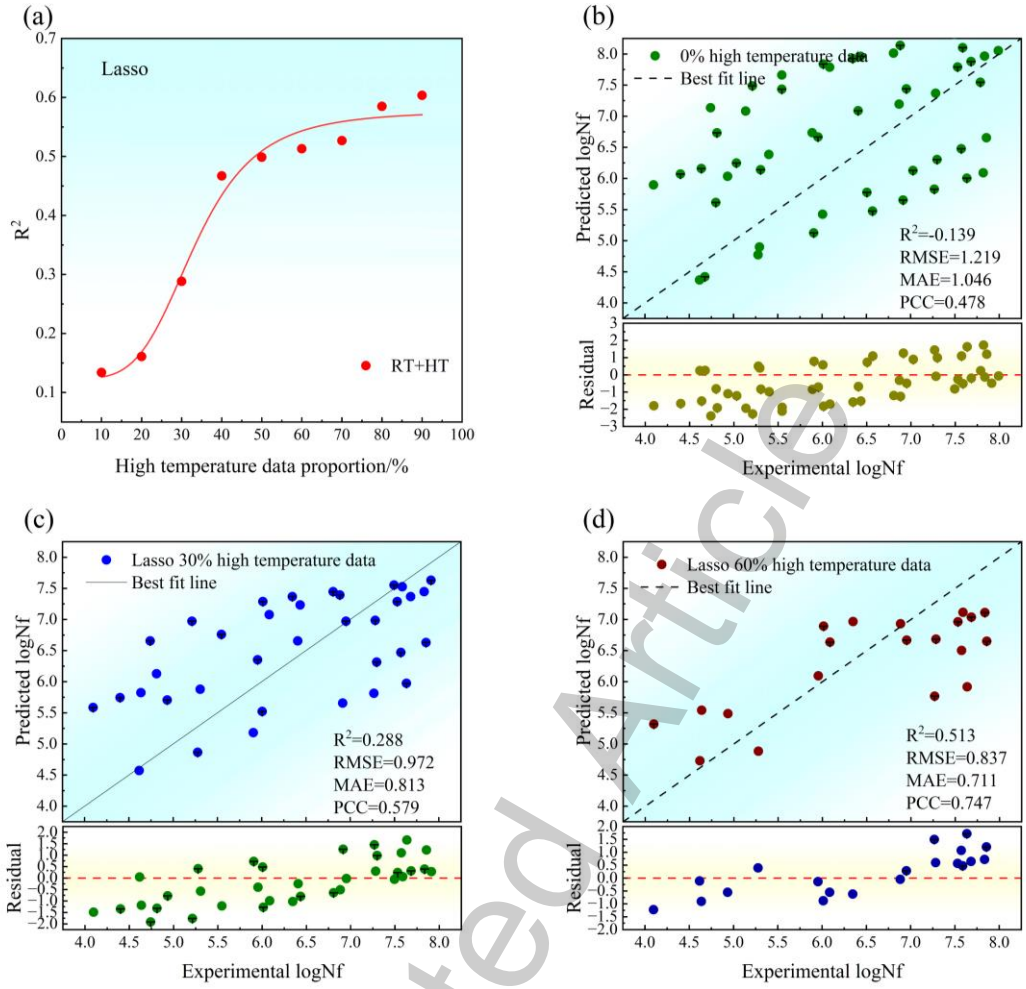

Fig. S8. Transfer learning prediction results of the Lasso model from room temperature to high temperature. (a) R^2 values of the model under different proportions of high-temperature data. (b–d) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0 %, 30 %, and 60 %, respectively.

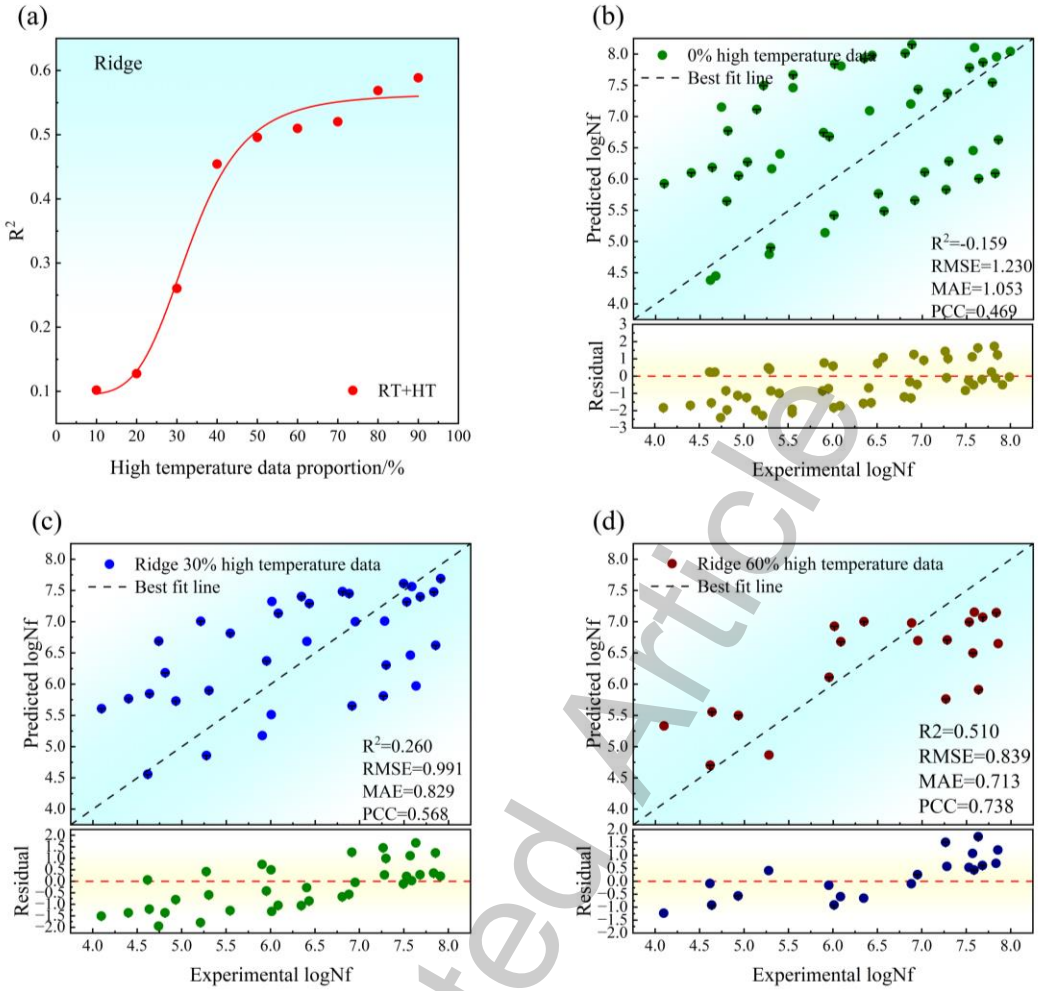

Fig. S9. Transfer learning prediction results of the Ridge model from room temperature to high temperature. (a) R^2 values of the model under different proportions of high-temperature data. (b–d) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0 %, 30 %, and 60 %, respectively.

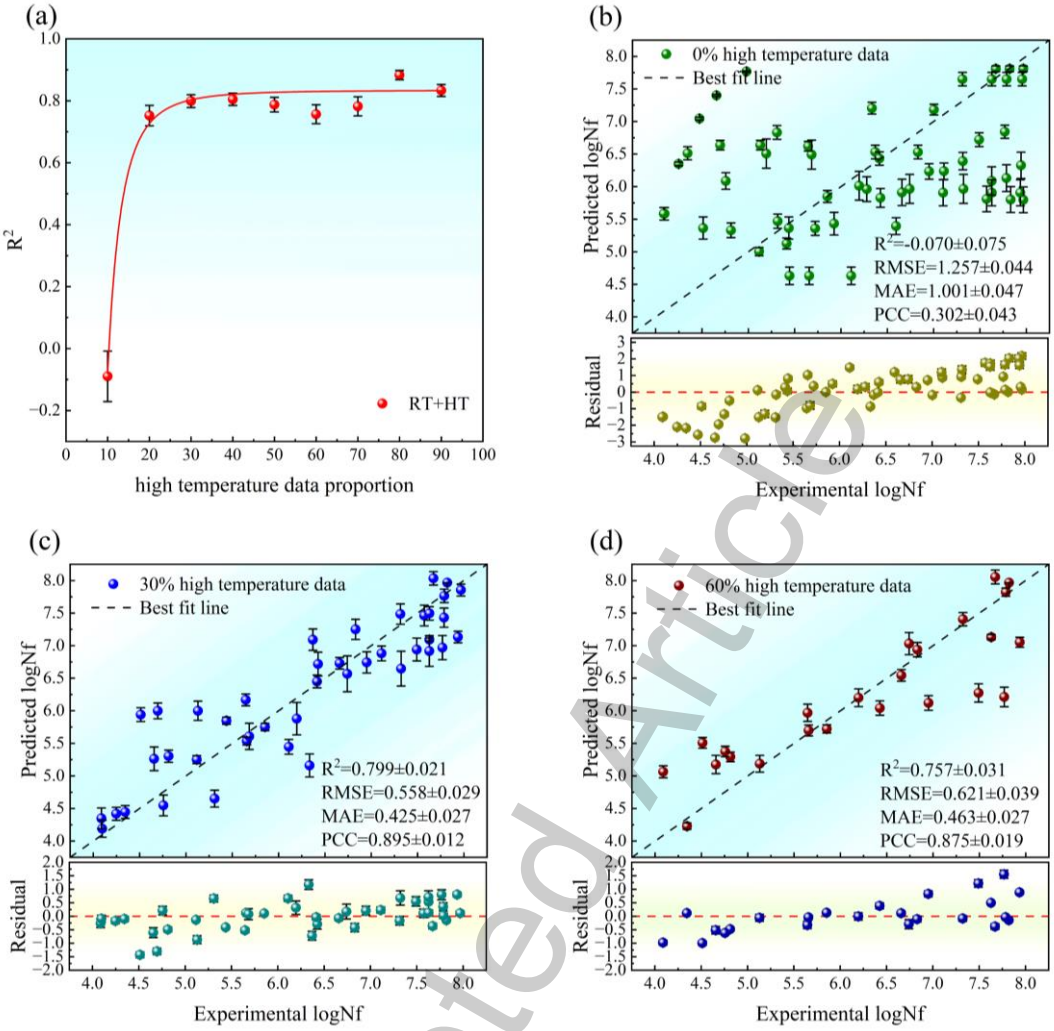

Fig. S10. Transfer learning prediction results of the GBDT model from room temperature to high temperature of 500 ℃. (a) R^2 values of the model under different proportions of high-temperature data. (b-d) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0 %, 30 %, and 60 %, respectively.

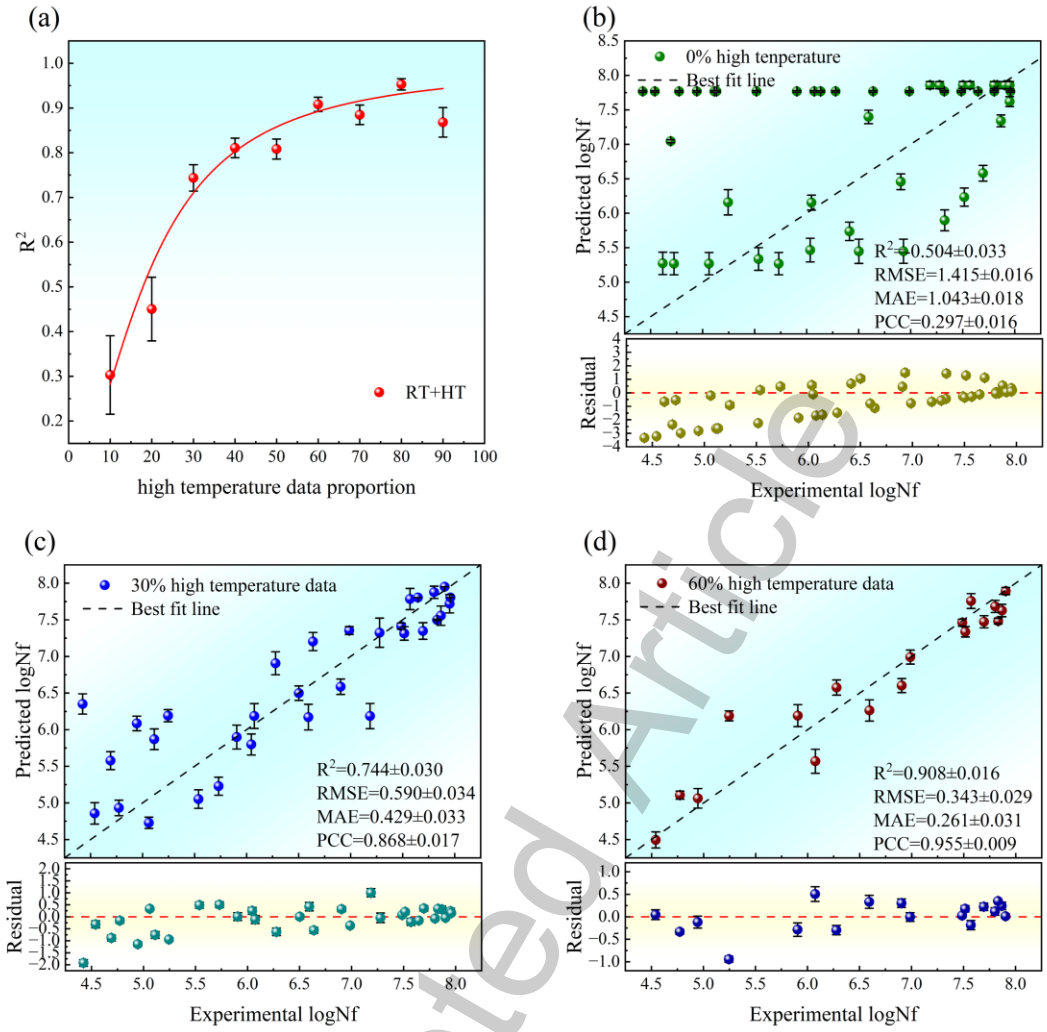

Fig. S11. Transfer learning prediction results of the GBDT model from room temperature to high temperature of 600 ℃. (a) R^2 values of the model under different proportions of high-temperature data. (b-d) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0 %, 30 %, and 60 %, respectively.