Research Article

A federated learning-driven data fusion strategy for the hardenability prediction of gear steel

Chunlei Shang[1,#], Tongbo Jiang[3,#], Lei Zhang[1,4], Hong-Hui Wu[1,2,3,*], Binbin Wang[1], Shuize Wang[1,2], Junheng Gao[1,2], Haitao Zhao[1,2], Chaolei Zhang[1,2], Xinping Mao[1,2]

[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 Steel Sustainable Technology, Liaoning Academy of Materials, Shenyang 110004, Liaoning, China.

[3]Institute of Materials Intelligent Technology, Liaoning Academy of Materials, Shenyang 110004, Liaoning, China.

[4]Angang Steel Research Institute Linggang Technology Center, Lingyuan 122500, Liaoning, China.

[*]**Correspondence to:** 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: hhwuaa@connect.ust.hk

How to cite this article: Shang, C.; Jiang, T.; Zhang, L.; Wu, H. H.; Wang, B.; Wang, S.; Gao, J.; Zhao, H.; Zhang, C.; Mao, X. A federated learning-driven data fusion strategy for the hardenability prediction of gear steel. *J Mater Inf*, 2026, *6*, [Accept]. http://dx.doi.org/10.20517/jmi.2026.26

Received: 7 May 2026 | **Revised:** 20 July 2026 | **Accepted:** 31 July 2026

Abstract

Hardenability is a critical indicator for evaluating the mechanical performance and service reliability of gear steels. However, conventional Jominy end-quench testing is labor-intensive and time-consuming, and data sharing among different companies is often restricted, which further complicates hardenability assessment. To address these challenges, federated learning–driven data fusion strategy incorporating a multi-regularized attention residual network model for hardenability prediction (MRAN-J9) is proposed. In this strategy, collaborative models are trained on heterogeneous data from multiple sources, improving predictive accuracy while preserving the privacy of each participant's raw data. Additionally, stable predictive performance is evaluated on a completely independent external validation dataset containing 755 samples (R^2 = 0.88, RMSE = 0.99 HRC), demonstrating the generalization capability and predictive stability. The results confirm the feasibility and effectiveness of federated learning for privacy-preserving multi-party collaborative modeling. Furthermore, integrating the MRAN-J9 model facilitates the effective exploitation of distributed multi-source data, providing a practical and reliable solution for hardenability prediction in complex industrial application scenarios.

Keywords: Federated learning, hardenability, Jominy end-quench test, machine learning, parameter optimization

INTRODUCTION

Hardenability is a critical parameter for evaluating the service performance of high-strength steel components such as gears and shafts[1]. It describes the ability of steel to develop uniform hardness and microstructural characteristics throughout the cross-section during quenching heat treatment, thereby directly influencing the mechanical properties and service life of the components[2,3]. At present, steel hardenability is most commonly assessed using the Jominy end-quench test. Originally proposed by Jominy and Boegehold[4], this standardized method characterizes hardenability by measuring the hardness distribution along the length of a cylindrical specimen subjected to water cooling at one end, yielding the corresponding end-quench curve[5,6]. Despite its widespread adoption and high degree of standardization, the Jominy test is prone to experimental uncertainties arising from oxide scale formation, microstructural inhomogeneity, and operator-dependent variability. Moreover, the testing procedure is relatively labor-intensive and time-consuming, limiting its suitability for rapid and cost-effective evaluation of steel hardenability in practical applications[7,8].

To enable rapid evaluation of the hardenability of steel materials, the literature has proposed a wide range of empirical and semi-empirical models, along with various predictive equations[9-11]. However, these approaches generally depend on restrictive assumptions and simplified parameter relationships, which limit their ability to capture the complex, nonlinear interactions inherent in multi-component steel systems. In recent years, the rapid advancement of machine learning techniques has led to significant progress in data-driven prediction and design of materials properties[12-14]. Among these approaches, federated learning has emerged as a promising distributed learning framework that effectively balances predictive performance and data privacy, making it particularly attractive for industrial applications[15,16]. To enable accurate prediction of defects in sheet metal forming processes, Dib *et al.*[17] proposed a federated learning-based framework integrated with digital envelopes, which effectively predicted defects in sheet metal forming. To accurately predict the welding quality of friction stir welding, Chakraborty *et al.*[18] developed a federated learning-based monitoring system that effectively predicted weld quality. Similarly, Wu *et al.*[19] proposed a three-layer federated learning framework for the collaborative training of deep learning models, which is successfully applied to quality defect detection in construction projects. These findings demonstrate that federated learning enables collaborative modeling across companies and production lines while preserving data security, and that its range of applications in industrial environments is steadily expanding.

In this work, experimental hardenability data of gear steels from four steel companies are collected, and the J9 value is used as the target performance indicator for modeling and prediction. The methodological advance of this work does not lie in introducing a new elementary neural-network operator. Instead, it develops a federated regression framework for hardenability, in which the network architecture, regularization strategy, and server aggregation rule are jointly designed for gear steel data from multiple sources. Specifically, efficient channel attention (ECA) is used to reweight coupled alloying element channels, preactivation residual blocks are employed to stabilize learning from moderately sized tabular datasets, dual-noise regularization is calibrated to composition measurement fluctuations and optimization uncertainty, and hierarchical feature aggregation is introduced to combine element-wise sensitivity with nonlinear alloying interactions. In addition, an adaptive sample loss aggregation rule is designed to mitigate non-independent and identically distributed (non-IID) client drift, extending beyond conventional aggregation relying on sample size. This strategy enables the security of the clients' raw data and improves the robustness of J9 hardenability prediction under heterogeneous industrial data distributions.

METHODS

Data source and processing

The data utilized in this work are collected from the actual industrial production lines of four major steelmaking enterprises, resulting in a final dataset of 4,244 gear steel samples. Specifically, these four enterprises contributed 1,058, 1,016, 1,100, and 1,070 samples, respectively. Each sample comprises 25 chemical composition variables and the J9. In the standard Jominy end-quench test, J9 represents the Rockwell Hardness (HRC) measured at a specific distance of 9 mm from the quenched end of the specimen. All experiments are conducted under uniform testing conditions and parameter settings, including a sample diameter of 25 mm, a length of 100 mm, a boss diameter of 30 mm with a length of 3 mm, and an austenitizing treatment at 950 °C for 1 h prior to end quenching. The experimental parameters are consistent across all samples; therefore, heat treatment process variables are not included as input features in the machine learning models. The analysis instead focused exclusively on the intrinsic relationship between chemical composition and hardenability.

Machine learning modeling and validation

This study employs a range of machine learning regression algorithms to predict J9 hardness values, including the K-Nearest Neighbors Regressor (KNN), Linear Regression (LR), Ridge Regression (RR), Support Vector Machine Regressor (SVM), Gradient Boosting Decision Tree (GBDT), Extreme Gradient Boosting (XGB), Random Forest (RF), and Light Gradient Boosting Machine (LGBM). The predictive performance of these models is systematically evaluated using the coefficient of determination (R^2) and the root mean square error (RMSE). The R^2 metric quantifies the goodness of fit between predicted and experimental values, providing a quantitative measure of the model's ability to explain data variability [20,21]. In contrast, RMSE reflects the overall magnitude of prediction errors, with smaller values indicating higher predictive accuracy. The mathematical definitions of R^2 and RMSE are provided below to ensure a consistent, standardized, and comparable evaluation of model performance in predicting J9 hardness values.

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

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

where y_i, $\hat{y}_i$, and $\bar{y}$ are the actual value, predicted value, and the mean of actual values.

To address composition-based hardenability prediction using industrial data, a multi-regularized attention residual network, termed MRAN-J9, is developed in this work. The model is designed not to introduce new elementary neural network operators, but to construct a regression architecture tailored to this specific task. Specifically, this architecture aims to capture interactions among alloying elements, improve robustness to industrial measurement fluctuations, and stabilize learning under non-IID data distributions across multiple clients. The MRAN-J9 architecture consists of four functional stages: input regularization, attention-enhanced residual feature extraction, hierarchical feature aggregation, and regression output.

In the input and feature-extraction stages, the normalized chemical-composition vector is first processed by GELU-based nonlinear transformations and an ECA layer. GELU provides smooth nonlinear mapping and preserves weak but informative composition responses. The GELU function is defined as follows:

$$GELU(x) = 0.5x(1 + \mathrm{erf}\,({}^{x}/_{\sqrt{2}})) \tag{3}$$

where $\mathrm{GELU}(x)$ denotes the output of the Gaussian error linear unit activation function, x represents the input feature of the neuron, and $\mathrm{erf}\,(.)$ is the Gaussian error function. The coefficient 0.5 and the normalization term $\sqrt{2}$ originate from the cumulative distribution formulation of the standard normal distribution. Meanwhile, ECA adaptively reweights alloying-element channels to capture cross-channel dependencies.

This design is physically relevant because J9 hardenability is controlled by coupled alloying effects rather than by individual elements alone. Pre-activation residual blocks are then used to improve gradient propagation and stabilize local training, which is important for moderate-size client datasets under federated learning.

To further improve generalization, dual-noise regularization is introduced. This dual-noise design is functionally complementary and dataset-tailored: the Gaussian noise imposes input-space robustness against actual spectrometer errors ($\sigma = 0.003$, ~0.1% perturbation of the standardized feature scale), while the gradient noise acts on optimization dynamics to prevent convergence to sharp minima, a critical safeguard for small-sample industrial data (~10^3 per client).

Federated learning strategy

Federated learning is a privacy-preserving distributed machine learning paradigm developed to address the data silo challenge arising from the difficulty of centrally sharing data across multiple sources. The concept is first introduced by Google in 2016 to enable on-device model training and updates on Android systems without transmitting raw data, thereby safeguarding user privacy[22,23]. For this work, the goal is for four clients to collaboratively train a global regression model without sharing data, achieving an accurate prediction of the target variable J9. The core modules of the federated learning strategy include data preprocessing and federated standardization, FedProx local training, adaptive global aggregation, and model evaluation and early stopping. Specifically, for each client, local training is terminated early if the local validation loss did not decrease for 10 consecutive local epochs. At the server side, the global aggregation process is terminated if the global validation loss showed no improvement for 20 consecutive global communication rounds. The maximum number of global communication rounds is uniformly set to 100 as the hard upper bound. During the execution, the global early stopping criterion is not activated, so the model training proceeded through the entire 100 rounds. The detailed training configurations and hyperparameter settings, including federated training schemes, optimization parameters, FedProx coefficient, adaptive aggregation coefficient, loss function settings, and regularization strategies, are summarized in Supplementary Table 1 of the Supplementary Material. The total loss in FedProx training consists of two parts:

regression loss (Huber Loss) and a proximal term penalty. The formula is as follows:

$$\vartheta_{local} = \vartheta_{huber}(y, \hat{y}) + \frac{\mu}{2} \sum_{p \epsilon \theta} \left\| p - p_{global} \right\|_2^2 \tag{4}$$

where ϑ_{local} denotes the total local loss of an individual client; $\vartheta_{hubert}(y, \hat{y})$ represents the Huber loss; μ is the proximal-term penalty coefficient used to control the strength of the constraint between local parameters and global parameters, with $\mu \in (0,1]$; p denotes the p_th parameter of the client's local model, and p_{global} denotes the p_th parameter of the global model at the current round; $\|\cdot\|_2^2$ denotes the squared L_2 norm, which is used to compute the Euclidean distance between parameter vectors. The Huber loss is a piecewise smooth loss function: it degenerates to the mean squared error (MSE when the error is small and transitions to the mean absolute error (MAE) when the error is large, thereby combining the gradient stability of MSE with the robustness to outliers of MAE. Its piecewise definition is given in Eq. (5).

$$\vartheta_{huber}(y, \hat{y}) = \begin{cases} \frac{1}{2}(y - \hat{y})^2, \ |y - \hat{y}| \le \delta \\ \delta \left(|y - \hat{y}| - \frac{\delta}{2} \right), \ otherwise \end{cases} \tag{5}$$

where y denotes the ground-truth label of the sample, $\hat{y}$ denotes the predicted label produced by the model, δ is the threshold parameter of the Huber loss that determines the transition between its piecewise definitions and is typically set adaptively according to the data distribution (in this study, δ=1.0), and $|\cdot|$ denotes the absolute value operator.

In conventional FedAvg, the contribution of each client is usually determined only by its local sample size. This strategy may be insufficient for gear steel data from multiple plants because different companies may have distinct composition ranges, hardness distributions, and measurement fluctuations. A client with a larger dataset may dominate the global update even when its local distribution is biased, whereas a client with fewer samples but more stable local optimization may be underrepresented. Therefore, an adaptive sample-loss aggregation strategy is introduced in this work. The sample-size term reflects the statistical contribution of each client, while the loss-based term reflects the reliability of the current local update. By jointly considering these two factors, the proposed aggregation strategy dynamically adjusts client contributions during federated training and helps mitigate client drift under non-IID industrial data distributions.

To replace the traditional weighting method based solely on the number of samples, an adaptive weighting strategy is designed by jointly considering the number of samples and the loss. The combined weight calculation formula is as follows:

$$\omega_k = (1-\alpha)\omega_k^{sample} + \alpha\omega_k^{loss} \tag{6}$$

where ω_k denotes the final adaptive weight of the *K-th* client; α is the weight allocation coefficient used to balance the relative contributions of the sample size and the loss value, with $\alpha \in (0,1)$; ω_k^{sample} denotes the normalized weight based on the sample size; and ω_k^{loss} denotes the normalized weight based on the local loss.

The formula for calculating the sample size normalized weight is as follows:

$$\omega_k^{sample} = \frac{N_k}{\sum_{m=1}^{K} N_m} \tag{7}$$

where N_k denotes the number of samples held by the *K-th* client; *K* denotes the total number of clients participating in federated training (in this study, $K = 4$); and $\sum_{m=1}^{K} N_m$ denotes the total number of samples across all clients.

The formula for calculating the normalized loss weight is:

$$\omega_k^{loss} = \frac{\vartheta_{max} - \vartheta_k}{\sum_{m=1}^{K}(\vartheta_{max} - \vartheta_m) + \epsilon} \tag{8}$$

where $\vartheta_{\max}$ denotes the maximum local training loss among all clients in the current round; ϑ_k denotes the local training loss of the *K-th* client; and $\sum_{m=1}^{K}(\vartheta_{\max} - \vartheta_m)$ denotes the sum of the differences between the maximum loss and the losses of all clients.

Based on adaptive weights, the local model parameters of each client are weighted and summed to obtain the updated global model parameters. The formula is as follows:

$$p_{global}^{new} = \sum_{k=1}^{K} \omega_k \cdot p_k \qquad (9)$$

where p_{global}^{new} denotes the p_th parameter of the updated global model; p_k denotes the p_th parameter of the local model of the *K-th* client; and $\sum_{k=1}^{K} \omega_k \cdot p_k$ represents the weighted sum of the local parameters of all clients using the adaptive weights.

RESULTS AND DISCUSSION

The overall workflow of the proposed federated learning strategy is illustrated in Figure 1. In the traditional approach, models are trained independently by each client using only local datasets; however, predictive performance is frequently hindered by limited data volumes and heterogeneous sample distributions. To overcome these constraints, the federated learning strategy is employed, where local models are uploaded to a central server for parameter aggregation. This process updates the global model, which is subsequently redistributed to the clients to achieve significantly enhanced predictive accuracy. Crucially, this strategy precludes direct data exchange between clients, thereby ensuring the security of the clients' raw data.

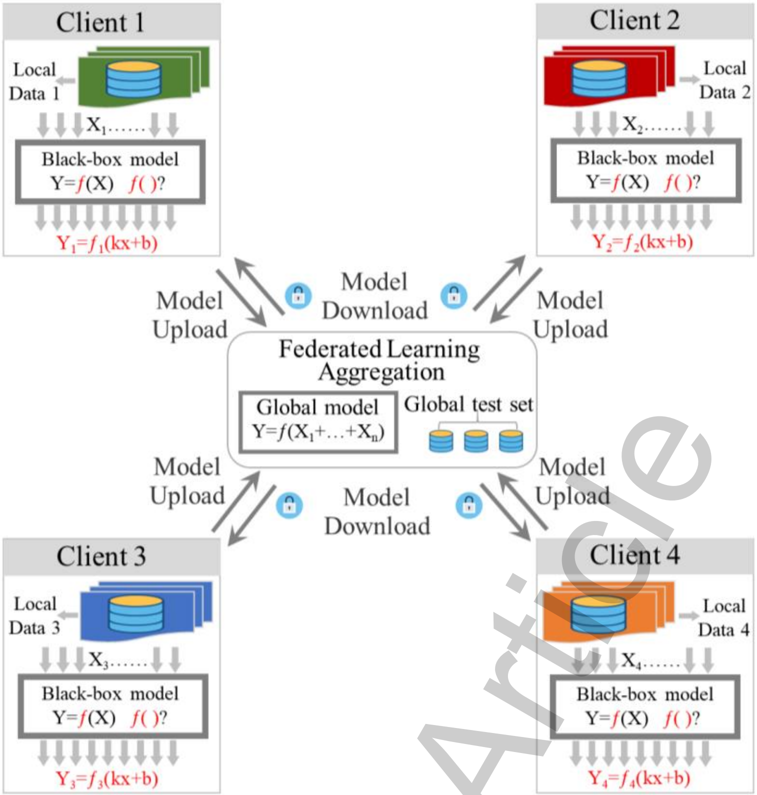

Figure 1. Server-client federated learning strategy and its training procedure.

The dataset is divided into training and testing sets at a ratio of 8:2, and the data distributions of the four clients are illustrated in Figure 2. The results indicate that for each client, both the training and testing sets generally follow an approximately normal distribution. This suggests that the data-splitting process preserves good statistical consistency, which is conducive to stable model training and reliable performance evaluation. Nevertheless, notable differences remain among the four clients in terms of data range and distribution characteristics. Specifically, the J9 hardness values are primarily distributed within the range of 28-44 HRC; however, substantial variations are observed across clients with respect to mean values, distribution widths, and sample densities. These discrepancies reflect the typical characteristics of multi-source heterogeneous data. The detailed statistical evaluation of each dataset is presented in Supplementary Table 2 of the Supplementary Materials.

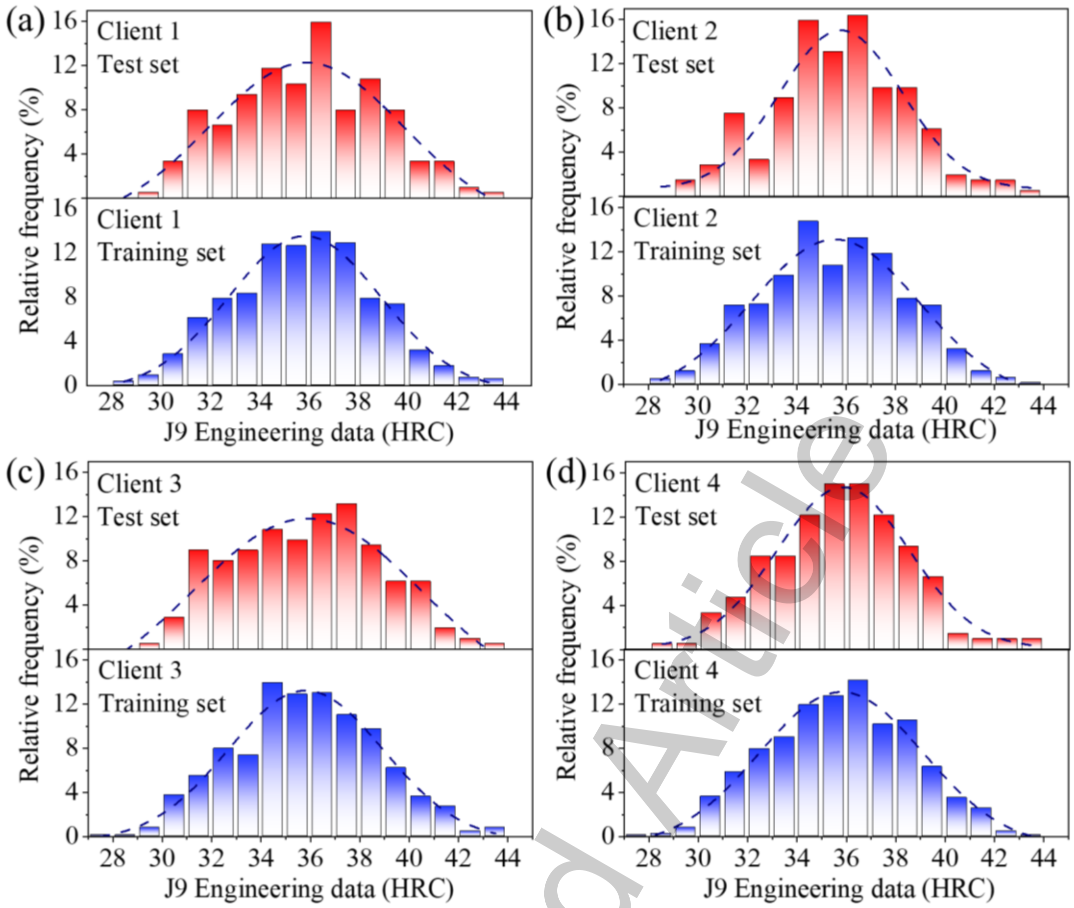

Figure 2. Data distribution of training and testing sets for machine learning from four clients. (a), (b), (c), and (d) correspond to clients 1, 2, 3, and 4, respectively.

To better accommodate the modeling requirements of multi-source heterogeneous data within a federated learning strategy, a novel MRAN-J9 model is proposed in this study. The overall network architecture is illustrated in Figure 3a. The model is composed of four principal components: the Input Layer, the Residual Block Area, the Feature Aggregation Layer, and the Output Layer. In the Input Layer, multi-dimensional feature data are received and normalized. Within the Residual Block Area, residual connections are introduced to effectively mitigate the gradient vanishing problem during deep network training, thereby enhancing the model's capability to capture complex nonlinear feature representations. The Feature Aggregation Layer is employed to fuse features extracted from different network layers, further strengthening the comprehensive modeling of salient information. Finally, the Output Layer is used to generate the prediction of the target variable. Compared with conventional machine learning models, the MRAN-J9 model demonstrates clear advantages in feature representation capacity and generalization performance, making it more suitable for meeting the accuracy and stability requirements of practical production-line applications.

During the model evaluation phase, 20% of the data from each client is reserved as a testing set to assess predictive performance. A comparison of the prediction performance of the MRAN-J9 model and eight conventional machine learning models across datasets from four clients is presented in Figure 3b. The eight traditional machine learning models are trained independently using local data. The results indicate that MRAN-J9 consistently achieves superior predictive performance across all clients, which outperforms the remaining eight machine learning models, as reflected by the highest R^2 and the lowest RMSE. These results demonstrate enhanced fitting capability and more stable predictive behavior (Client 1: $R^2 = 0.79$, RMSE = 1.32 HRC; Client 2: $R^2 = 0.68$, RMSE = 1.55 HRC; Client 3: $R^2 = 0.68$, RMSE = 1.67 HRC; Client 4: $R^2 = 0.72$, RMSE = 1.41 HRC). The detailed predictive results for the other eight machine learning models are summarized in Supplementary Table 3 of the Supplementary Material. Collectively, the results indicate that the MRAN-J9 model exhibits significant advantages in multi-client, multi-distribution data environments and is effective in improving overall prediction performance. Besides traditional machine learning, representative deep learning backbones (1D-CNN, MLP, and SE-MLP) trained solely on local data are also evaluated. As detailed in Supplementary Table 4, the MRAN-J9 model consistently surpasses these modern deep learning baselines even before federated aggregation, and the gap is substantially widened when the federated strategy is applied, confirming the necessity of both the customized architecture and the collaborative training paradigm.

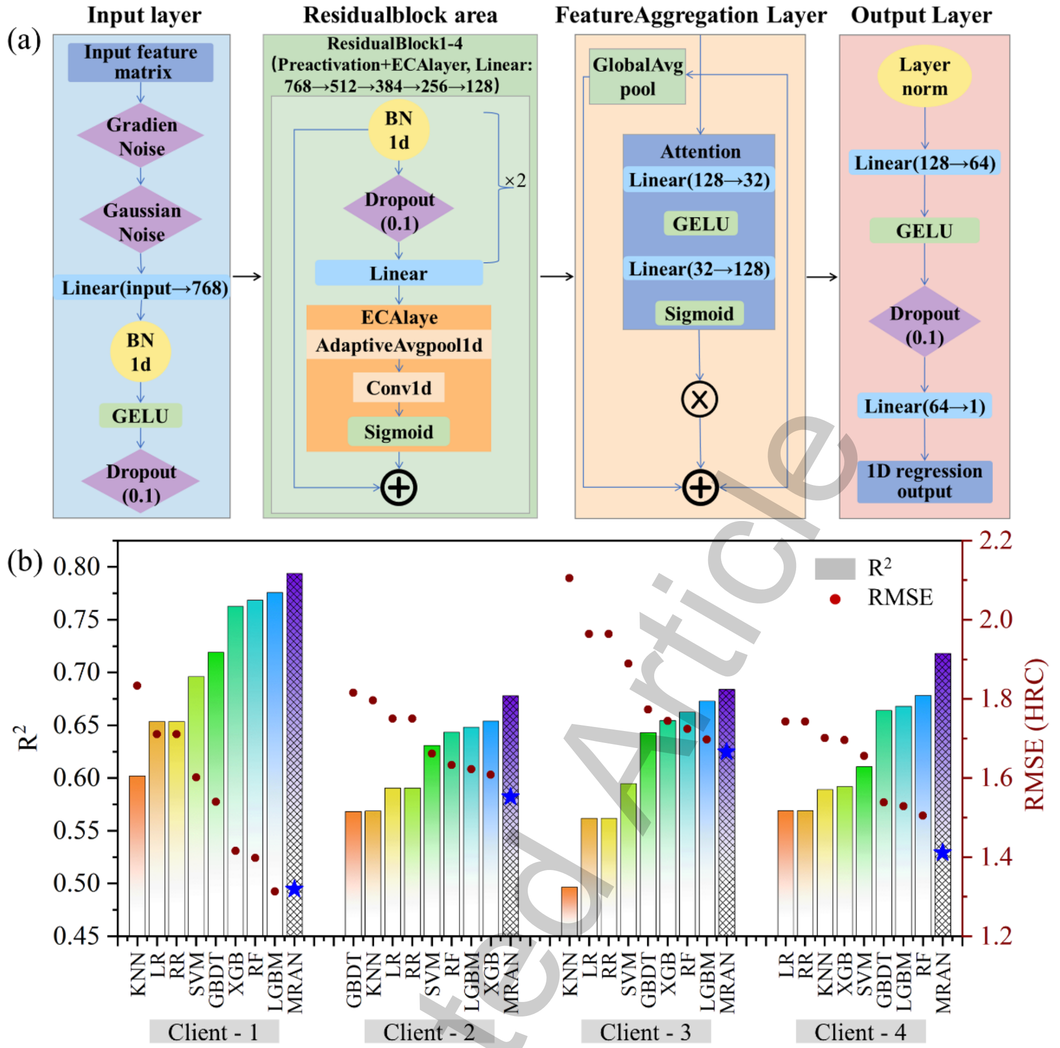

Figure 3. Model architecture and model prediction results. (a) MRAN-J9 architecture. (b) Model prediction results from multiple machine learning models across four clients.

The parameter optimization process of the MRAN-J9 model for four clients within the federated learning strategy, together with the predictive performance of the corresponding optimal-parameter models, are presented in Figure 4. In each lobal communication rounds, the central server aggregates client-updated model parameters to update a global model, which is subsequently broadcast back to the clients to enable collaborative training while keeping raw data local. Each client then performs iterative local updates of the MRAN-J9 parameters on its own dataset, progressively converging to a client-specific optimum. For client 1, Figure 4a1 shows the optimization results over 100 global communication rounds, Figure 4a2 reports the prediction performance of the optimal model (R^2 = 0.88, RMSE = 1.03 HRC), and Figure 4a3 depicts the frequency distribution of the predicted values. Figs. 4(b–d) present the parameter optimization process and the corresponding prediction performance for clients 2-4 (Client 2: $R^2 = 0.80$, RMSE = 1.25 HRC; Client 3: $R^2 = 0.84$, RMSE = 1.19 HRC; Client 4: $R^2 = 0.83$, RMSE = 1.34 HRC), respectively. Furthermore, the proportions of prediction errors falling within the ±2 HRC range for the four clients are 94.39%, 92.99%, 89.72%, and 94.85%, respectively. Across all clients, predictions agree closely with the ground truth, with most points clustering around the 1:1 line, indicating that MRAN-J9 delivers consistent and robust predictive performance under the federated learning strategy. Beyond predictive accuracy, a SHAP feature-importance analysis is conducted on the datasets of all four clients. The results reveal that the top six most influential features (Mn, Cr, C, Ti, Si, and B) are identical across all clients. Notably, these elements are widely recognized in physical metallurgy for their significant influence on steel hardenability. This consistency effectively validates the physical reliability of the model. The detailed results are presented in Supplementary Figure 1 of the Supplementary Materials.

To further clarify the source of the performance improvement, two levels of controlled comparisons are conducted. First, architectural ablation methods are performed under a purely local setting. The residual connection, ECA attention, noise regularization, and feature aggregation layer are removed individually while the remaining training settings are kept unchanged. As summarized in Supplementary Table 5 of the Supplementary Material, the full MRAN-J9 model achieved an average R^2 of 0.72 across the four clients. After removing the residual connection, ECA attention, noise regularization, and feature aggregation layer, the average R^2 decreased to 0.66, 0.67, 0.67, and 0.66, respectively. These results indicate that each architectural module contributes to the predictive capability of MRAN-J9 under the purely local setting.

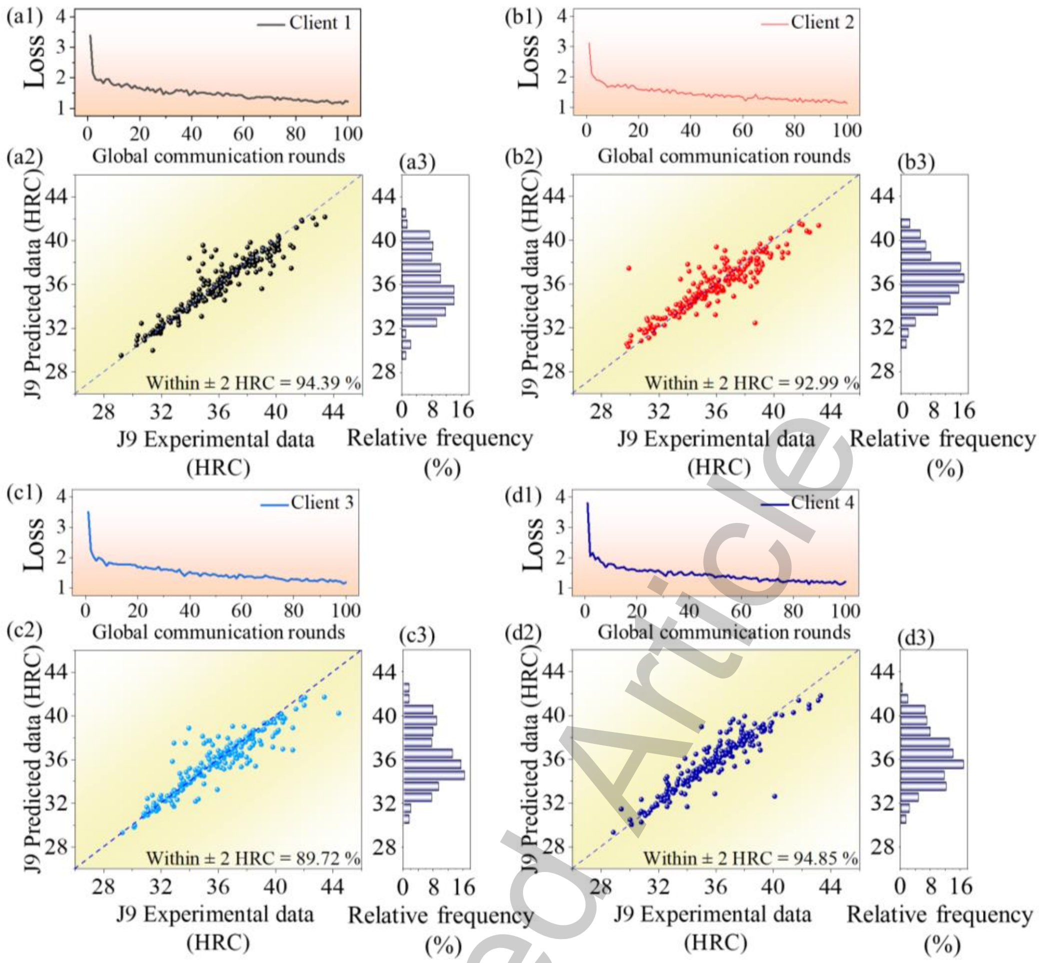

Figure 4. Parameter optimization of the federated learning strategy and prediction performance of the model with optimal parameters. (a1) Optimization results over 100 global communication rounds for Client 1; (a2) prediction performance of the optimal-parameter model; (a3) histogram of the frequency distribution of predicted values. (b-d) Optimization and prediction results for Clients 2, 3, and 4, respectively.

To further distinguish the contribution of the federated aggregation strategy from the backbone architecture, benchmarks for both FedAvg and FedProx are additionally established under the identical MRAN-J9 network. The detailed ablation comparison is summarized in Supplementary Table 6. The results indicate that the adaptive weighting and robust preprocessing, rather than the proximal term alone, are the primary drivers of the performance enhancement, confirming the rationality of the federated learning design.

The results presented in Figure 4 demonstrate that the MRAN-J9 model trained under the federated learning strategy can effectively predict the test-set data of all four clients, as illustrated in Figure 5a. High predictive accuracy and good stability are maintained across different client data distributions. To further evaluate the generalization capability of the MRAN-J9 model, an additional validation dataset containing 755 samples is constructed. This validation dataset is derived from a different source than the original datasets of the four clients. The prediction results obtained on the validation dataset are shown in Figure 5b, where the R^2 of 0.88 and the RMSE of 0.99 HRC are achieved. Additionally, the model demonstrates excellent practical accuracy, with 93.38% of the errors strictly confined within the ±2 HRC tolerance. These results indicate that the MRAN-J9 model is able to maintain robust predictive performance when applied to new data not used during training. Moreover, the findings demonstrate that the federated learning strategy effectively mitigates overfitting while fully leveraging multi-source data information, thereby significantly enhancing the model's generalization capability and practical applicability.

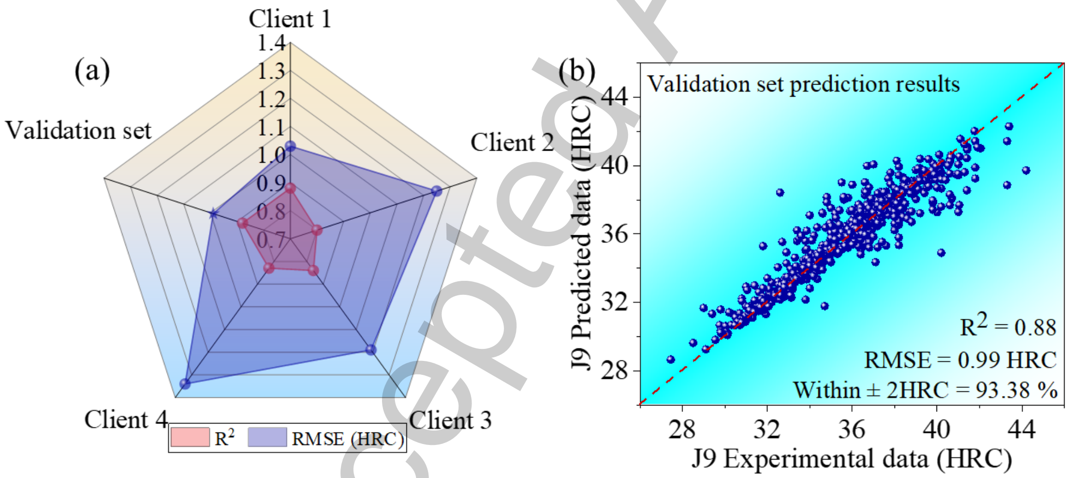

Figure 5. Federated learning validation set results. (a) The model prediction results for four clients and the results for the validation set. (b) Comparison of experimental and predicted values in the validation set.

In traditional centralized machine learning, distributed data are aggregated at a central server for model training[24-26]. In contrast, federated learning enables the construction of shared models among multiple participants in a secure, efficient, and regulation-compliant manner without requiring data to leave their respective domains[27,28]. As demonstrated by the results in Figs. 4 and 5, the MRAN-J9 model trained under the federated learning strategy is able to effectively adapt to heterogeneous data distributions across different clients while exhibiting strong generalization performance on an external independent validation dataset. These findings further validate the advantages of federated learning for collaborative modeling involving multi-source heterogeneous data.

With the continued advancement of federated learning theory and techniques, its application scope has been progressively extended to a wide range of domains, including intelligent healthcare[29], recommendation systems, smart cities[30], finance and insurance, and edge computing[31]. In response to the increasing complexity of application requirements, federated learning algorithms and system architectures are being actively investigated to achieve improved trade-offs among data security, communication efficiency, and model performance. Through the facilitation of cross-institutional collaboration via secure connection platforms, federated learning is driving data collaboration paradigms toward more legal, compliant, and sustainable practices. In this context, the MRAN-J9 model proposed in this study provides effective technical support for the deployment of federated learning in industrial application scenarios.

CONCLUSION

In this work, an MRAN-J9 model suitable for the federated learning strategy is proposed and systematically validated using data from four clients. The results demonstrate that the MRAN-J9 model trained under the federated learning strategy effectively adapts to the heterogeneous data distribution characteristics of different clients, achieving superior predictive performance on each client's test set when compared with non-federated learning modeling approaches. Furthermore, high predictive accuracy is maintained on a completely independent external validation dataset (R^2 = 0.88, RMSE = 0.99 HRC), thereby demonstrating the model's strong generalization capability and stability. These findings further confirm the feasibility and effectiveness of federated learning in enabling collaborative modeling across multiple data owners while protecting client raw data and complying with regulatory requirements. By integrating MRAN-J9 into the federated learning strategy, latent feature information from multi-source data is more fully exploited, and the limitations of single-source modeling are alleviated. This integrated approach offers a new paradigm for researching the properties of gear steel, facilitating broader application deployment across environments involving multiple enterprises and production lines.

DECLARATIONS

Acknowledgements

The computing work is supported by USTB MatCom of Beijing Advanced Innovation Center for Materials Genome Engineering.

Authors' contributions

Writing - original draft, software, methodology, formal analysis, data curation: Chunlei Shang, Tongbo Jiang

Writing - review and editing, supervision, project administration, investigation: Hong-Hui Wu, Binbin Wang

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

Visualization, resources, formal analysis, conceptualization: Haitao Zhao, Chaolei Zhang, Lei Zhang

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

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

AI and AI-assisted tools statement

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.

Conflicts of interest

Hong-Hui Wu is a Youth Editorial Board Member of the journal *Journal of Materials Informatics*, 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.

Copyright

REFERENCES

1. Erden M A, Aydın F. Wear and mechanical properties of carburized AISI 8620 steel produced by powder metallurgy. International Journal of Minerals, Metallurgy and Materials, 2021, 28: 430-439. DOI: 10.1007/s12613-020-2046-8
2. Xue H, Peng W, Yu L, et al. Effect of hardenability on microstructure and property of low alloy abrasion-resistant steel. Materials Science and Engineering: A, 2020, 793: 139901. DOI: 10.1016/j.msea.2020.139901
3. Di Schino, Emilio Di Nunzio, Maria Cabrera. Effect of quenching & partitioning process on a low carbon steel. Advanced Materials Letters, 2017, 8: 641-644. DOI: 10.5185/amlett.2017.1487
4. Jominy W, Boegehold A. A hardenability test for carburizing steel. Trans. ASM, 1938, 26: 574-606.
5. Tenaglia N E, Boeri R E, Massone J M, et al. Assessment of the austemperability of high-silicon cast steels through Jominy hardenability tests. Materials Science and Technology, 2018, 34: 1990-2000. DOI: 10.1080/02670836.2018.1507124
6. Cakir M, Özsoy A. Investigation of the correlation between thermal properties and hardenability of Jominy bars quenched with air–water mixture for AISI 1050 steel. Materials & Design, 2011, 32: 3099-3105. DOI: 10.1016/j.matdes.2010.12.035
7. Newkirk J W, MacKenzie D S. The Jominy end quench for light-weight alloy development. Journal of Materials Engineering and Performance, 2000, 9: 408-415. DOI: 10.1361/105994900770345809
8. Hömberg D. A numerical simulation of the Jominy end-quench test. Acta Materialia,

1996, 44: 4375-4385. DOI: 10.1016/1359-6454(96)00084-5

9. Yu B H. Nonlinear equations for calculating hardenability and mechanical property. Iron and Steel(China), 1985, 20: 40-49.

10. Jin M, Lian J, Jiang Z. New method for prediction of Jominy curve of structural steel. Acta Metallurgica Sinica(China), 2006, 42: 405-410.

11. Geng X, Cheng Z, Wang S, et al. A data-driven machine learning approach to predict the hardenability curve of boron steels and assist alloy design. Journal of Materials Science, 2022, 57: 10755-10768. DOI: 10.1007/s10853-022-07132-9

12. Shang C L, Zhu D X, Wu H H, et al. A quantitative relation for the ductile-brittle transition temperature in pipeline steel. Scripta Materialia, 2024, 244: 116023. DOI: 10.1016/j.scriptamat.2024.116023

13. Xian Y, Dang P, Tian Y, et al. Compositional design of multicomponent alloys using reinforcement learning. Acta Materialia, 2024, 274: 120017. DOI: 10.1016/j.actamat.2024.120017

14. Song L, Wang C, Li Y, et al. Predicting stacking fault energy in austenitic stainless steels via physical metallurgy-based machine learning approaches. Journal of Materials Informatics, 2025, 5: 2. DOI: 10.20517/jmi.2024.70

15. Jiang J, Hu L, Hu C, et al. Bacombo—bandwidth-aware decentralized federated learning. Electronics, 2020, 9: 440. DOI: 10.3390/electronics9030440

16. Hao, M, Li, H, Luo, X, et al. Efficient and privacy-enhanced federated learning for industrial artificial intelligence. IEEE Transactions on Industrial Informatics, 2019, 16(10), 6532-6542. DOI: 10.1109/TII.2019.2945367

17. da Silveira Dib MA, Prates P, Ribeiro B. SecFL–Secure Federated Learning Framework for predicting defects in sheet metal forming under variability. Expert Systems with Applications, 2024, 235: 121139. DOI: 10.1016/j.eswa.2023.121139

18. Chakraborty S, Guha S, Mishra D, et al. Federated learning for weld quality prediction. Journal of Dynamics, Monitoring and Diagnostics, 2024, 3: 237-245. DOI: 10.37965/jdmd.2024.529

19. Wu H T, Li H, Chi H L, et al. A hierarchical federated learning framework for collaborative quality defect inspection in construction. Engineering Applications of Artificial Intelligence, 2024, 133: 108218. DOI: 10.1016/j.engappai.2024.108218

20. 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. Corrosion

Communications, 2026, 21: 60-72. DOI: 10.1016/j.corcom.2025.12.002

21. Xu M, Xing S, Hong J, et al. A hybrid deep learning model for robust and data-efficient lithium-ion battery remaining useful life prediction. Journal of Materials Informatics, 2026, 6: 26. DOI: 10.20517/jmi.2025.80
22. McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data. Artificial intelligence and statistics, 2017, 54: 1273-1282. DOI: 10.48550/arXiv.1602.05629
23. Yang Q, Liu Y, Chen T, et al. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology, 2019, 10: 1-19. DOI: 10.1145/3298981
24. Hao C, Kuai P, Duan J, et al. Machine-learning-enabled composition–process codesign of heat-resistant cast aluminum alloys with superior elevated-temperature strength. Journal of Materials Informatics, 2026, 6: 38. DOI: 10.20517/jmi.2026.20
25. Guo S, Yu J, Liu X, et al. A predicting model for properties of steel using the industrial big data based on machine learningJ.. Computational Materials Science, 2019, 160: 95-104. DOI: 10.1016/j.commatsci.2018.12.056
26. Shen C, Wang C, Wei X, et al. Physical metallurgy-guided machine learning and artificial intelligent design of ultrahigh-strength stainless steelJ.. Acta Materialia, 2019, 179: 201-214. DOI: 10.1016/j.actamat.2019.08.033
27. Li A, Zhang L, Wang J, et al. Privacy-preserving efficient federated-learning model debugging. IEEE Transactions on Parallel and Distributed Systems, 2021, 33: 2291-2303. DOI: 10.1109/TPDS.2021.3137321
28. Li, T, Sahu, A K, Talwalkar, A, et al. Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 2020, 37(3), 50-60. DOI: 10.1109/MSP.2020.2975749
29. Rieke, N, Hancox, J, Li, W, et al. The future of digital health with federated learning. npj Digital Medicine, 2020, 3(1), 1-7. DOI: 10.1038/s41746-020-00323-1
30. Zheng, Z, Zhou, Y, Sun, Y, et al. Applications of federated learning in smart cities: recent advances, taxonomy, and open challenges. Connection Science, 2022, 34(1), 1-28. DOI: 10.48550/arXiv.2102.01375
31. Zhou H, Yang G, Dai H, et al. PFLF: Privacy-preserving federated learning framework for edge computing. IEEE Transactions on Information Forensics and Security, 2022, 17: 1905-1918. DOI: 10.1109/TIFS.2022.3174394