Supplementary Materials

Predicting stacking fault energy in austenitic stainless steels via physical metallurgy-based machine learning approaches

Longyu Song, Chenchong Wang, Yizhuang Li, Xiaolu Wei*

State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110819, Liaoning, China.

***Correspondence to:** Dr. Xiaolu Wei, State Key Laboratory of Rolling and Automation, Northeastern University, NO. 3-11, Wenhua Road, Heping District, Shenyang 110819, Liaoning, China. E-mail: weixl@smm.neu.edu.cn

This PDF file includes:

Professional Terminology Explanations

Supplementary Tables 1 to 2

Supplementary Figures. 1 to 3

Professional Terminology Explanations

The thermodynamic theoretical calculation of stacking fault energy is based on the thermodynamic model proposed by Olson-Cohen, which is a formal solution model. This model views the occurrence of stacking faults in crystals as a transformation between FCC (face-centered cubic) and HCP (hexagonal close-packed) structures, allowing for the calculation of energy changes per unit area. As shown in **Equation S1**, these three physical parameters are derived from the empirical formula used to calculate stacking fault energy. We have calculated the values of these PM parameters such as the phase transformation driving force (DF), and the Gibbs free energies of the FCC (G_{FCC}) and HCP (G_{HCP}) phases separately using the Thermo-Calc®. The empirical formulas for calculating the physical parameters related to the thermodynamic model of stacking fault energy are listed in **Supplementary Table 1**.

$$\gamma = 2\rho_A \triangle G^{\gamma\to\varepsilon} + 2\sigma^{\gamma\to\varepsilon} \quad (S1)$$

$$\triangle G^{\gamma\to\varepsilon} = \triangle G_{che}^{\gamma\to\varepsilon} + \triangle G_{mg}^{\gamma\to\varepsilon} + \triangle G_{seg}^{\gamma\to\varepsilon} \quad (S2)$$

$$\triangle G_{che}^{\gamma\to\varepsilon} = \sum_i X_i \triangle G_i^{\gamma\to\varepsilon} + \sum_{ij} X_i X_j \triangle \Omega_{ij}^{\gamma\to\varepsilon} \quad (S3)$$

$$\triangle G_{mg}^{\gamma\to\varepsilon} = \triangle G_m^{\varepsilon} - \triangle G_m^{\gamma} \quad (S4)$$

$$G_m^{\varphi} = RT \ln(\beta + 1) f(\tau) \quad (S5)$$

$$\sigma^{\gamma\to\varepsilon} = 10(2\rho_A(\triangle G_{che}^{\gamma\to\varepsilon} + \triangle G_{mg}^{\gamma\to\varepsilon}))^2 + 0.0095 \quad (S6)$$

Where ρ_A represents the molar surface density of the closest-packed plane in a face-centered cubic (FCC) crystal. $\triangle G_{che}^{\gamma\to\varepsilon}$ is the molar thermochemical free energy difference; $\triangle G_{mg}^{\gamma\to\varepsilon}$ is the magnetic free energy difference, which is related to the Néel temperatures of the respective phases; $\triangle G_{seg}^{\gamma\to\varepsilon}$ is the free energy difference arising from the Suzuki effect between the γ and ε phases, which is negligible at 300K due to its small magnitude. X_i is the molar fraction; $\triangle G_i^{\gamma\to\varepsilon}$ is the free energy change for the γ→ε transition in pure component i; and $\triangle \Omega_{ij}^{\gamma\to\varepsilon}$ represents the interaction energy between components i and j. The specific empirical formulas for these calculations are shown in Supplementary Table 1, where φ denotes either the γ or ε phase; R is the gas

constant; and $\tau = T/T_N$, where T is the room temperature and T_N is the antiferromagnetic transition temperature for the R phase. For both FCC and HCP structures, p = 0.28.

Supplementary Table 1. Empirical Formulas for Calculating Physical Parameters Related to the Thermodynamic Model of SFE

Parameter	Function (J /mol)
$\triangle G_{Fe}^{\gamma\to\varepsilon}$	-2243.38+4.309T
$\triangle G_{Cr}^{\gamma\to\varepsilon}$	1370-0.163T
$\triangle G_{Ni}^{\gamma\to\varepsilon}$	1046+1.255T
$\triangle G_{Mn}^{\gamma\to\varepsilon}$	-1000.00+1.123T
$\triangle G_{Si}^{\gamma\to\varepsilon}$	-560-8T
$\Omega_{FeCr}^{\gamma\to\varepsilon}$	2095
$\Omega_{FeNi}^{\gamma\to\varepsilon}$	2095
$\Omega_{FeMn}^{\gamma\to\varepsilon}$	2873-717 $(X_{Fe} - X_{Mn})$
$\Omega_{FeSi}^{\gamma\to\varepsilon}$	2850 + 3520 $(X_{Fe} - X_{Si})$
f(τ)	when τ<1, $1 - [\frac{79\tau^{-1}}{140p} + \frac{474}{497}\left(\frac{1}{p} - 1\right)\left(\frac{\tau^3}{6} + \frac{\tau^9}{135} + \frac{\tau^{15}}{600}\right)]/A$
	when τ>1, $-\left(\frac{\tau^{-5}}{10} + \frac{\tau^{-15}}{315} + \frac{\tau^{-25}}{1500}\right)]/A$

The SHAP, which is based on game theory and local explanations, can explain the relationship between the inputs and targets using the Shapley value. The Shapley value can be expressed as **Equation S7**:

$$\varphi_i = \sum_{S\subseteq F\backslash\{i\}} \frac{|S|!\,(|F| - |S| - 1)}{|F|!}\left(\nu_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - \nu_S(x_S)\right) \quad (S7)$$

where φ_i is the Shapley value of the ith input value, S is the set of input values with the ith input value excluded, |S| is the magnitude of S (for example, S = x_1, x_2, . . . , x_{i-1}, x_{i+1}, . . . ,x_{n-1}, x_n and |S| = n − 1), F is the set of input values, and v(x) is calculated based on the marginal contribution of the input values.

The MDA, which is a feature selection method within the realm of machine learning, quantifies the importance of each input feature by measuring the decrease in

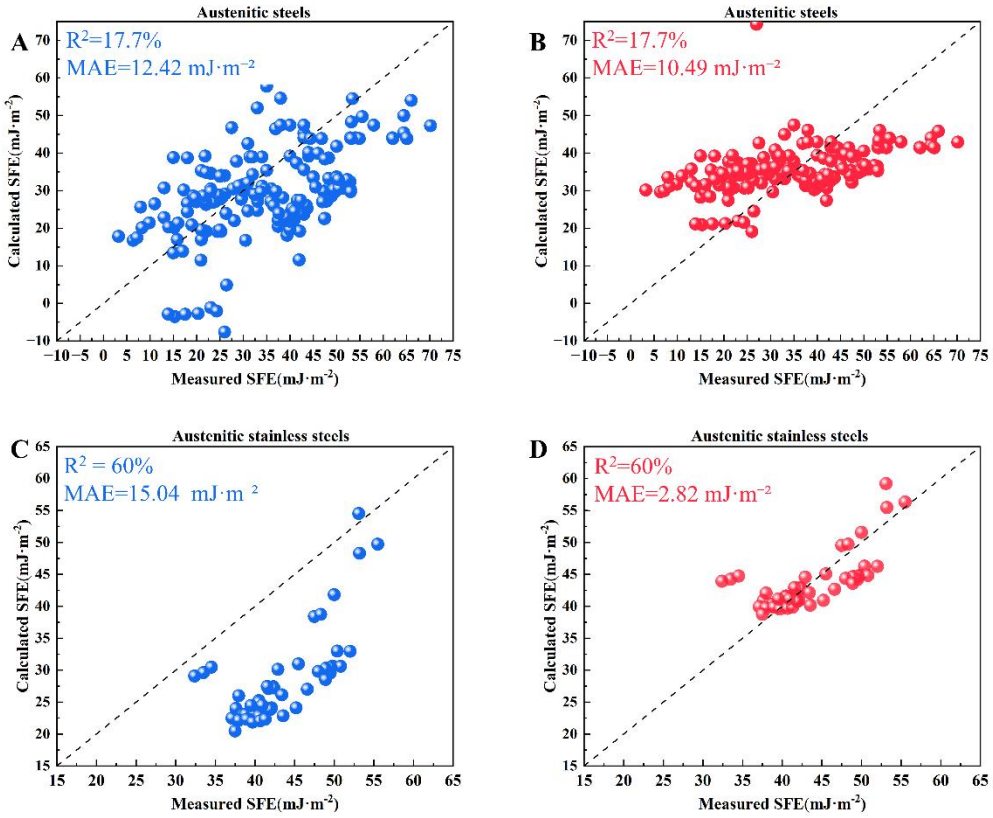
model accuracy, as shown in the **Equation S8**:

$$\mathrm{MDA} = \frac{1}{\mathrm{T}}\sum_{\mathrm{t}}^{\mathrm{T}}\left[\frac{1}{|\mathrm{D_t}|}\left(\sum_{\mathrm{X_i^j}\in\mathrm{D_t^j}}\sum_{\mathrm{k}}\left(\mathrm{R_k}\left(\mathrm{X_i^j}-\mathrm{y_i^j}\right)\right)^2-\sum_{\mathrm{X_i}\in\mathrm{D_t}}\sum_{\mathrm{k}}\left(\mathrm{R_k}(\mathrm{X_i})-\mathrm{y_i^k}\right)^2\right)\right] \quad (S8)$$

where T is the total number of decision trees in the Random Forest model, $\mathrm{D_t}$ means the out-of-bag (OOB) sample set for the *t*th decision tree. The OOB data is used to estimate the model's accuracy without using the same data for both training and validation, thus providing an unbiased estimate of the model's performance. $\mathrm{X_i^j}$ is the *i* th sample with the *j* th feature permuted. $\mathrm{D_t^j}$ is the OOB sample set for the *t*th decision tree with the *j*th feature permuted. $\mathrm{R_k}(\mathrm{X_i})$ is the predicted output of the Random Forest model for the *i*th sample on the *k*th target variable, and $\mathrm{y_i^k}$ is the actual *k*th target variable value for the *i*th sample.

Conventional SFE calculation results

Before using ML models, we attempted to calculate the SFE using the thermodynamic model. To assess the validity of the empirical formulas in calculating SFE, we conducted SFE calculations for 188 alloy compositions in the database and conducted a detailed comparative analysis of these calculated results with experimentally obtained SFE values. **Supplementary Figure 1** shows the distribution of calculated and actual SFE values for the alloy compositions in the database, with the effectiveness evaluated using R^2 and MAE indicators. The comparison revealed significant differences between the calculated values obtained using the thermodynamic empirical formulas and the experimentally measured values. Specifically, the R^2 of the model was only 17.7%, indicating a very poor model fit. Additionally, the MAE was as high as 12.42 $mJ \cdot m^{-2}$, further demonstrating the significant challenges in predicting SFE for alloy compositions in the thermodynamic database. Given the rigorous and complex process of determining interfacial energy, we attempted to evaluate a subset of 54 austenitic stainless steel compositions in the database separately, as shown in **Supplementary Figure 1C**. After this segmentation, the R^2 value improved significantly to 60%. However, the MAE increased to 15.04 $mJ \cdot m^{-2}$, suggesting that although the model fit improved to some extent, there was an overall numerical shift between the predicted and actual values.



Supplementary Figure 1. The distribution of calculated and actual SFE values for the alloy compositions in the database. A: austenitic steels before fit; B: austenitic steels after fit; C: austenitic stainless steels before fit; D: austenitic stainless steels after fit.

Taking into account the fact that the two key variables are subject to some error due to the empirical formulae, we decided to introduce a coefficient (as in Equation S9) in front of each variable in order to analyze the subsequent fitting work.

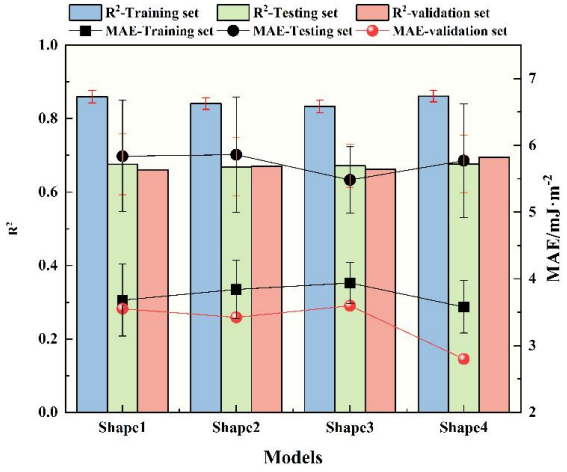
$$\gamma = \alpha * 2\rho_A \triangle G^{\gamma\to\varepsilon} + \beta * 2\sigma^{\gamma\to\varepsilon} \quad (S9)$$

When fitting the austenitic steel compositions, by determining a set of parameters (α= 0.4335, β= 1.6129), we find that the MAE value decreases to 10.49 mJ·m⁻², while for the austenitic stainless steels, it decreases significantly from 15.04 mJ·m⁻² to 2.82 mJ·m⁻², but the fitting was not as effective as expected. There are still challenges in realizing accurate prediction of the SFE in a wider range of compositions.

Detailed results of the SFE modeling approach

In the process of constructing a Convolutional Neural Network (CNN) model, a systematic and rigorous methodology was followed to determine the sizes of convolutional layers and fully connected layers. Here are the detailed steps: Firstly, to explore the optimal model structure, we conducted a series of parameter tuning experiments. These experiments involved altering key parameters such as the size and number of convolution kernels in the convolutional layers, as well as the number of neurons in the fully connected layers. To evaluate model performance, we adopted R^2 and MAE as metrics to help us compare the pros and cons of different model structures.

As shown in **Supplementary Figure 2**, after trying various parameter configurations, we found that introducing PM parameters into the fully connected layers did not significantly improve model performance. The role of these PM parameters was difficult to clearly identify, and their effect on enhancing prediction performance was minimal. This finding prompted us to recognize that the current introduction strategy had limited effectiveness in improving model performance and motivated us to consider how to reduce computational complexity while maintaining model performance. **Supplementary Table 2** lists the corresponding architectural details for the model shapes depicted in Supplementary Figure 2. Based on the best training results, we gradually adjusted the sizes of the model's structural layers, aiming to reduce computational complexity while ensuring performance, so that the model could be trained within a reasonable timeframe.

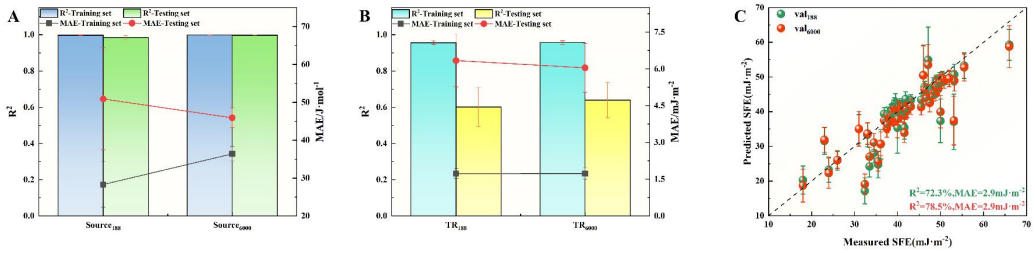


Supplementary Figure 2. The R^2 and MAE of models under different model shapes.

Supplementary Table 2. The CNN model architecture details of layer (type)

Shape1	Shape2	Shape3	Shape4
conv2d_1 (Conv2D): (4, 4, 8)	conv2d_1 (Conv2D): (4, 4, 16)	conv2d_1 (Conv2D): (4, 4, 8)	conv2d_1 (Conv2D): (4, 4, 8)
conv2d_1 (Conv2D): (4, 4, 32)	conv2d_1 (Conv2D): (4, 4, 32)	conv2d_1 (Conv2D): (4, 4, 16)	conv2d_1 (Conv2D): (4, 4, 16)
flatten (Flatten):2336	flatten (Flatten):4640	flatten (Flatten):1168	flatten (Flatten):1168
dense_1 (Dense): 256	dense_1 (Dense): 256	dense_2 (Dense): 128	dense_1 (Dense): 64
dense_2 (Dense): 128	dense_2 (Dense): 128	dense_3 (Dense): 8	dense_2 (Dense): 8
dense_3 (Dense): 8	dense_3 (Dense): 8		

The results of the transfer learning model after expanding the database are shown in **Supplementary Figure 3**.



Supplementary Figure 3. Comparing the results of the transfer training model after expanding the data volume. A: source model; B: transfer model; C: distribution of predicted and measured values of SFE for austenitic stainless steels.