INDIAN INSTITUTE OF TECHNOLOGY, KANPUR
AI-Assisted Prediction of Stacking Fault Energy in HSLA Steels for Enhanced Mechanical Performance
MSE497- UNDERGRADUATE PROJECT-II

Nikunj Muchhal - 220713
Dhruv Gupta - 220362
Dhreeaj Yadav - 220356
Under the guidance of:
Prof. Amarendra K Singh &
Prof. Soumya Sridhar
Department of Metallurgical and Materials Engineering
IIT Kanpur

Abstract

Stacking fault energy (SFE) plays a crucial role in shaping the mechanical behavior of high-strength low-alloy (HSLA) steels. It directly influences how dislocations move, how the material hardens under stress, and the onset of deformation mechanisms like twinning and transformation. In this project, we aimed to predict SFE based purely on alloy compositions by combining domain knowledge from metallurgy with modern machine learning techniques. Using a curated dataset from the literature—primarily from the detailed study by Wang and Xiong (2020)—and enhancing it with data generated from CALPHAD-based simulations and synthetic augmentation, we built and trained models like Support Vector Machines (SVM) and Random Forests (RF). Both models achieved strong predictive performance and revealed meaningful insights into how elements like Mn, Cr, Ni, Si, and C affect SFE. The results not only matched known metallurgical trends but also demonstrated the potential of machine learning to accelerate alloy design and reduce dependence on traditional experimentation.

1 Introduction

Stacking fault energy (SFE) is one of those concepts in material science that doesn't get as much attention as it deserves, but when we started looking into it, we realized how crucial it is. SFE plays a central role in determining how a material behaves under stress, especially when we're talking about things like strength and toughness. To put it simply, SFE is the energy required to create a fault in a material's atomic structure, and it can have a huge impact on how a material performs in real-world conditions. In this paper, we're focusing on how we navigated throughout the project of predicting SFE energies for different compositions using prediction modelling and ML. Through our research, we aim to look into the mechanisms that link the impact of composition over SFE. The goal isn't just academic; it's about understanding how the atomic properties of materials can help improve the safety and performance of everyday technologies.

2 Objective

- Develop predictive models for SFE based on alloy composition.
- Identify influential elements impacting SFE.
- Validate model predictions against metallurgical theory.
- Support alloy design through predictive analytics.

3 Methodology

3.1 Literature-Based Data Compilation

A thorough review of peer-reviewed literature was used to carefully select the foundational dataset for this investigation, with a particular focus on Wang and Xiong's (2020) work. More than 300 data points about austenitic steels, including chemical compositions and associated SFE values, were obtained from their thorough analysis. Understanding the intricate relationship between alloying elements and SFE was made possible thanks in large part to this dataset. Key takeaways from their study include:

- **Elemental Influence:** Nickel (Ni) and Iron (Fe) exhibited a moderate monotonic effect on SFE, while other elements like Manganese (Mn),

Chromium (Cr), and Silicon (Si) showed non-linear and composition-dependent influences.

- **Data Variability:** The dataset encompassed a wide range of compositions, ensuring a diverse representation of austenitic steels.

Incorporating this data provided a solid foundation for training and validating machine learning models, ensuring they capture the nuanced effects of various alloying elements on SFE.

3.2 Computational Data Augmentation

To supplement the literature-derived data, computational methods were employed to generate additional data points:

- **Thermo-Calc & CALPHAD Simulations:** Utilizing the CALPHAD (Calculation of Phase Diagrams) approach, simulations were conducted to predict SFE values for hypothetical alloy compositions. This method allowed exploration beyond experimentally available data, providing insights into unexplored compositional spaces.

  The generated data points were used at certain points to accommodate the need of the literature but were not primarily used as it affected the Machine learning model curve

3.3 Data Preprocessing and Refinement

The initial dataset, comprising 474 entries, had multiple refinements based on our requirements:

3.4 Phase 1

- **Temperature Normalization:** Entries lacking temperature data were excluded, resulting in a refined dataset of 428 entries.
- **Inital Minor Element Filtering:** Data points with minor elements (¡0.05 wt%) were scrutinized, and those with negligible influence on SFE were removed, yielding 360 high-quality data points.
- **Phase:** Initially the data points included multiple phases without addressing the fact that it could affect our output

3.5 Phase 2

- **Improvements :** The Temperature Normalization was found correct, but eliminating the data points with a lesser composition led to anomalies in trends so we used data points with significant impact even with lower composition were included
- **Phase Consistency :** To remove the confusion because of multiple phases, we finally structured our study to Austenitic Steels only.

3.6 Dataset Splitting for Model Training

For effective model training and validation:

- **Training Set:** 80% of the data (288 entries) were allocated for training the machine learning models.
- **Test Set:** The remaining 20% (72 entries) were reserved for evaluating model performance on unseen data, ensuring the models' generalizability.

This strategic split facilitated strong model development, allowing for accurate prediction of SFE based on alloy compositions.

3.7 Benefits of Data Modeling

- **Normalization of Weight Percentages:** To maintain consistency throughout the data set, weight percentages of various alloying elements were normalized. This ensured that all elements were represented on a common scale, which is essential for the training of machine learning models that rely on numerical inputs. The normalization process minimizes bias that could arise from disproportionate element compositions.
- **Handling of Zero-Composition Values:** In several instances, zero-composition values were observed for certain elements. These entries were handled with care to avoid misinterpretation. Zero values could indicate either the absence of an element or that it was below the detection limits of the measurement method. This distinction was crucial for ensuring the integrity of the dataset and preventing errors in model predictions.

4 Machine Learning Model and Outputs

In this study, various regression models were applied to predict stacking fault energy (SFE) from material composition and structure. These included linear regression for baseline performance, kernel-based models like Support Vector Regressor (SVR) and Kernel Ridge for capturing non-linear relationships, and tree-based models such as Decision Tree, Random Forest, and Extra Trees Regressors for handling feature interactions.
Ensemble boosting techniques like Gradient Boosting and HistGradient Boosting were also used to iteratively improve predictions, along with K-Nearest Neighbors (KNN), a distance-based method relying on nearby data points.
Among all models, the best performances were observed with Extra Trees Regressor, Random Forest Regressor, and Gradient Boosting Regressor. These models provided high accuracy and generalization ability, making them the most suitable choices for SFE prediction.

4.1 Pearson Correlation Matrix :

The Pearson Correlation Matrix shows how strongly each element is linearly correlated with SFE (Stacking Fault Energy) and with each other. A value close to +1 indicates a strong positive correlation, -1 indicates a strong negative correlation, and 0 indicates no correlation. The top 5 features that showed the strongest correlation with SFE are Ni, Fe, Mo, Cr, and Mn.

4.2 SHAP Analysis:

SHAP (SHapley Additive exPlanations) analysis helps us understand how much each feature contributes to a model's prediction. It assigns an importance value (positive or negative) to every feature for each individual prediction, based on game theory concepts, making the model's decisions more transparent and explainable
The SHAP analysis shows that Mn, Si, P, Mo, and N are the most influential features affecting the model's prediction of stacking fault energy (SFE). Higher Mn content tends to increase the SFE, while higher Si and P generally push the predictions lower. The waterfall plot for a specific prediction shows that Mn had the strongest positive contribution, while Ni and C contributed negatively. Overall, the model's predictions result from a balance between positive and negative influences of different alloying elements

Model	Training R^2	Test R^2
Linear Regression	0.52	0.73
Support Vector	0.64	0.76
Decision Tree	0.72	0.55
Random Forest	0.92	0.78
Extra Trees	0.99	0.79
KNN	0.99	0.75
Gradient Boosting	0.99	0.78

Table 1: Model R^2*comparison*

4.3 Extra Trees regression :

Extra Trees regression is a variant of Random Forest that introduces additional randomness by selecting split points completely randomly instead of finding optimal thresholds. Unlike Random Forest, which uses bootstrapped samples, Extra Trees typically builds each tree using the entire dataset. This increased randomization makes Extra Trees faster to train while often maintaining comparable accuracy. The algorithm generally exhibits lower variance than Random Forest, making it more robust against noise and outliers in the data, though sometimes at the cost of slightly higher bias. The Extra Trees Regressor showed excellent performance on the training set with an R^2 score of 0.9928 and very low error. On the test set, it achieved a good R^2 score of 0.7997, indicating strong but slightly reduced generalization. The true vs.

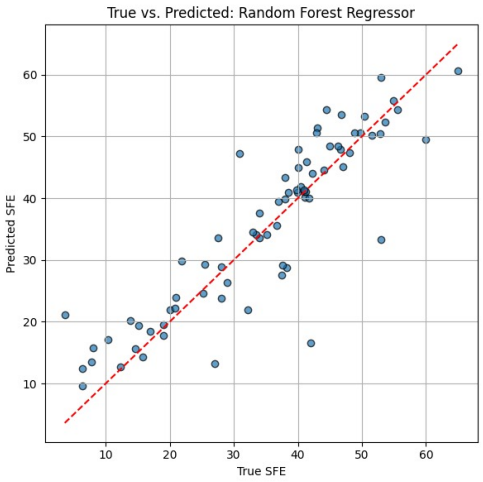

Figure 1: True v/s Predicted SFE

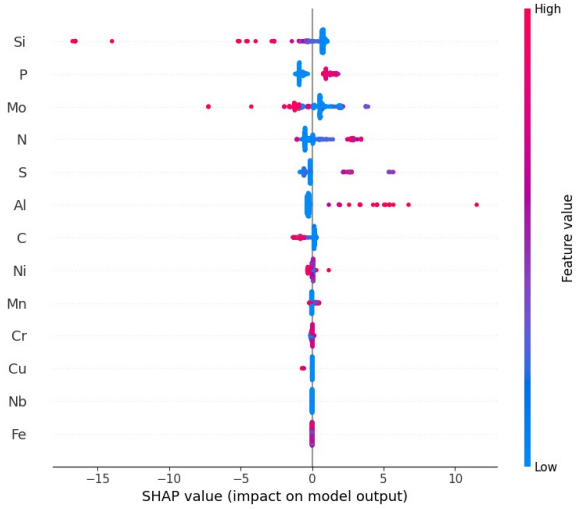

Figure 2: SHAP Analysis

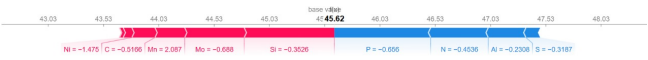

Figure 3: Key Features

predicted plot shows that most predictions closely follow the ideal trend.

5 Results and Discussion

5.1 Manganese (Mn) – Parabolic Influence

Manganese plays a dual role in influencing the stacking fault energy (SFE). At low to mid-range compositions (0–18 wt%), Mn stabilizes the austenitic phase, balancing the SFE and promoting Twinning-Induced Plasticity (TWIP). As the Mn content exceeds 18 wt%, it destabilizes the austenitic phase, causing SFE to drop again. This behavior forms a classic parabolic relationship. The minimum SFE near 25 wt% Mn corresponds to the transition zone between TWIP and Transformation-Induced Plasticity (TRIP). This observation is consistent with the findings of Wang and Xiong, and is accurately predicted by our machine learning models.

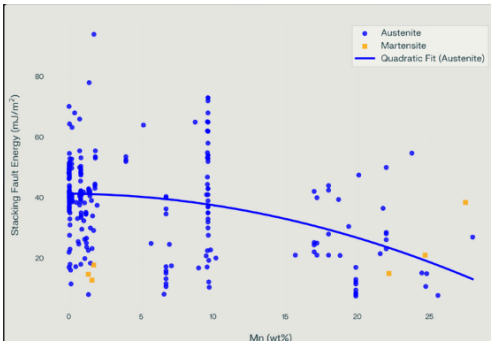

Figure 4: Mangnese showing the Parabolic Trend

5.2 Chromium (Cr) – Linearly Decreasing Trend

Chromium acts as a ferrite stabilizer, resulting in a linear decrease in SFE as the content of chromium increases. Beyond 20 wt%, this ferrite-stabilizing effect becomes more pronounced, and each 1 wt% of Cr reduces SFE by approximately 0.8–1.2 mJ/m². This behavior aligns with chromium's known role in promoting the FCC-to-BCC transformation, contributing to a decrease in stacking fault energy.

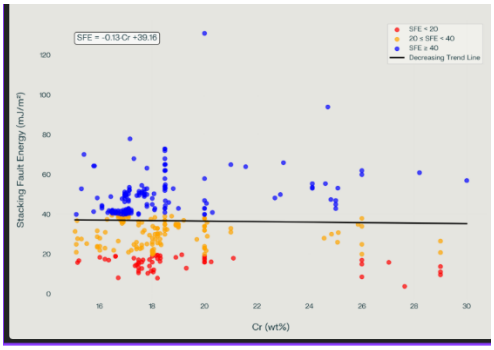

Figure 5: Decreasing trend of Chromium

5.3 Silicon (Si) – Sharp Linear Decrease

Silicon has a significant effect on SFE, reducing it by approximately 4.2 ± 0.3 mJ/m² for each 1 wt% increase in silicon. This reduction in SFE enhances the planarity of dislocations, which in turn promotes planar slip. Silicon's role is particularly beneficial in enhancing work hardening properties without inducing phase instability, which is crucial for improving material performance under stress.

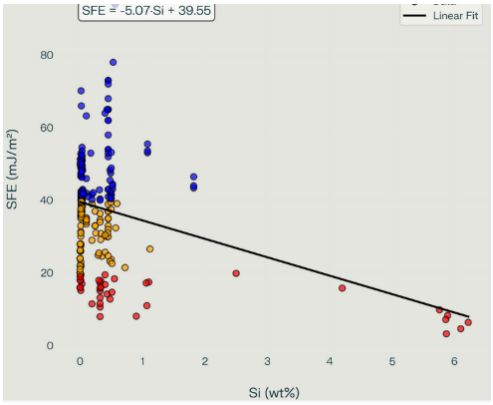

Figure 6: Decreasing trend by Silicon

5.4 Nickel (Ni) – Positive Linear Effect

Nickel is a potent austenite stabilizer, and its influence on SFE is positive and linear. For each 1 wt% increase in nickel content, SFE increases by approximately 1.5–2 mJ/m². This behavior is consistent with both the predictive models and findings in the literature, confirming that nickel's stabilizing effect on austenite contributes to an increase in stacking fault energy.

5.5 Carbon (C) – Mild to Neutral Influence

The influence of carbon on SFE is relatively mild, particularly within the ranges studied. While carbon plays a crucial role in stabilizing the austenitic phase, its effect on SFE is generally constant or slightly decreasing in some cases. This behavior can be attributed to the

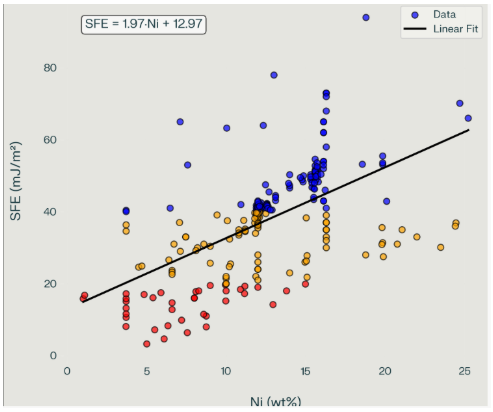

Figure 7: Increasing trend of Nickel

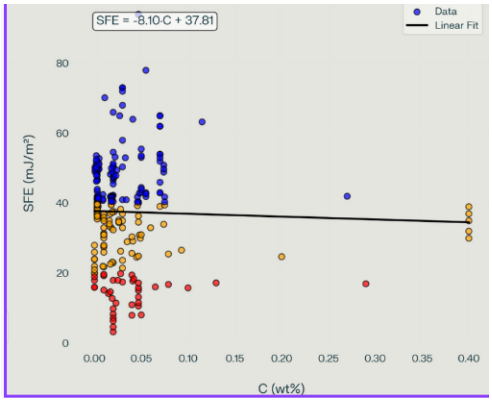

Figure 8: Decreasing trend in Carbon

interstitial nature of carbon, which interacts with substitutional solutes to affect phase stability rather than drastically altering SFE values.

6 Conclusion

This study successfully demonstrates how machine learning models can be used to predict stacking fault energy (SFE) based on alloy compositions. By combining data from literature, computational simulations, and synthetic data generation, we built accurate and interpretable models that shed light on the metallurgical behavior of high-strength low-alloy (HSLA) steels. The analysis highlighted several key trends, including:

- A parabolic relationship between Manganese (Mn) content and SFE.
- A linear decrease in SFE with increasing Chromium (Cr) and Silicon (Si).
- A positive linear correlation between Nickel (Ni) content and SFE.
- A neutral to mildly decreasing effect of Carbon (C) on SFE.

These observations align well with established metallurgical theories and previous research, while also offering fresh insights into how different alloying elements influence SFE. This work underlines the growing potential of AI in revolutionizing materials science—making it faster, more data-driven, and better equipped to tackle complex material challenges.

7 Learning & Outcomes

Working on this project gave us a deep appreciation for how machine learning can be applied to real-world material science problems. By combining metallurgical theory with data-driven modeling, We were able to explore the influence of elemental composition on stacking fault energy (SFE) in ways that would have been difficult using only experimental methods.

One of the biggest takeaways was learning how different alloying elements — especially manganese, chromium, nickel, silicon, and carbon — contribute to the mechanical behavior of HSLA steels. The machine learning models didn't just predict values; they helped us understand patterns that aligned with established metallurgical understanding, like the parabolic effect of manganese or the linear trends with chromium and silicon.

On the technical side, We developed a solid understanding of data preprocessing, model selection, and evaluation techniques. Implementing algorithms like Support Vector Machines (SVM) and Random Forests (RF), tuning their parameters, and interpreting the results helped build both confidence and clarity in working with machine learning tools.

8 Challenges Faced

Like any meaningful project, this one came with its share of challenges. One of the first and most persistent issues was the limited availability of high-composition manganese data. Since Mn plays a critical role in influencing stacking fault energy, this gap made it harder to fully capture and model its parabolic trend in the initial stages. We later addressed this by expanding our dataset, but the early models struggled to reflect Mn's true impact.

Another challenge was deciding how to handle very low-composition elements. Some entries included elements in trace amounts that either had negligible influence or introduced noise. Creating synthetic data to balance this out required careful judgment to avoid skewing the overall dataset or misrepresenting physical trends.

Data gathering itself was time-consuming, especially from literature where formats and reported variables varied. Standardizing this for analysis took significant effort.

Finally, identifying reliable trends and correlations wasn't straightforward. With only a few key features like composition and temperature available, the models had limited context. This made it tricky to explain certain deviations or justify correlations in a highly variable metallurgical landscape.

Despite these hurdles, each challenge led to a better understanding of both the subject and the modeling process, making the learning outcomes even better.

9 Acknowledgments

We would like to express our sincere gratitude to Prof. Amarendra K. Singh & Prof. Soumya Sridhar for their invaluable guidance throughout this research project.

Their continuous support and weekly insights were instrumental in shaping our understanding of the material behavior and the application of machine learning in metallurgy. We are especially grateful for their feedback, which helped us refine our approach and strengthen the overall quality of our analysis.

Additionally, we would like to acknowledge the authors of the papers and datasets we referenced, especially Wang and Xiong (2020), whose comprehensive analysis of austenitic steels provided a solid foundation for our research. The process of collecting and analyzing data from multiple sources helped us navigate through a wealth of literature, deepening our understanding of the material properties and advancing our research goals. This project has been an incredibly enriching learning experience, and we look forward to applying the knowledge gained here to future endeavors.

10 Bibliography & References

1. Lee, Y. K., Lee, S. J., & Han, J. (2016). Critical assessment 19: Stacking fault energies of austenitic steels. Materials Science and Technology, 32(1), 1–8. https://doi.org/10.1080/02670836.2015.1114252

2.Yonezawa, T., Suzuki, K., Ooki, S. et al. The Effect of Chemical Composition and Heat Treatment Conditions on Stacking Fault Energy for Fe-Cr-Ni Austenitic Stainless Steel. Metall Mater Trans A 44, 5884–5896 (2013). https://doi.org/10.1007/s11661-013-1943-0

3. Saeed-Akbari, A., Imlau, J., Prahl, U. et al. Derivation and Variation in Composition-Dependent Stacking Fault Energy Maps Based on Subregular Solution Model in High-Manganese Steels. Metall Mater Trans A 40, 3076–3090 (2009). https://doi.org/10.1007/s11661-009-0050-8

4. Schramm, R.E., Reed, R.P. Stacking fault energies of seven commercial austenitic stainless steels. Metall Trans A 6, 1345–1351 (1975). https://doi.org/10.1007/BF02641927