Thermodynamic Interpretation of Stress Corrosion Cracking in Austenitic Stainless Steels using Explainable AI

Han Gyeol Cho [a], Dayu Fajrul Falaakh [a], Chi Bum Bahn [a*]
[a] *School of Mechanical Engineering, Pusan National University, Busan, Republic of Korea*
[*] *Corresponding author: bahn@pusan.ac.kr*

Keywords : explainable artificial intelligence, stress corrosion cracking, crack growth rate, thermodynamic interpretation

1. Introduction

Stress Corrosion Cracking (SCC) poses a significant threat to the integrity of austenitic stainless steels in Light Water Reactor (LWR) environments. While machine learning (ML) has shown promise in predicting Crack Growth Rates (CGR), its 'black-box' nature limits the ability to identify physical causality. A recent study by Falaakh et al. [1] compiled a comprehensive CGR database and explored ML modeling with a primary focus on uncertainty quantification.

Unlike the previous work, the present study utilizes the established database to develop an independent, distinct CatBoost-based prediction model. The primary contribution of this research is extending the ML model to serve as a tool for understanding thermodynamic corrosion mechanisms. By integrating Explainable Artificial Intelligence (XAI), this study aims to demonstrate how data-driven models can identify underlying thermodynamic patterns - such as phase boundaries on E-pH diagrams and thermodynamic driving forces (ΔECP) - directly from complex empirical databases.

2. Methods

A comprehensive database of 978 CGR data points (BWR and PWR conditions) from the literature was utilized to train a CatBoost algorithm [2]. CatBoost was selected due to its robustness against data noise and efficient handling of categorical variables without loss of information. To interpret the model, Shapley Additive exPlanations (SHAP) values [3] were calculated. To quantify the combined thermodynamic influence of the water chemistry, the SHAP values of pH and ECP were summed and then standardized into Z-scores (Z_ϕ).

To evaluate the thermodynamic oxidation driving force across varying temperatures, ΔECP was introduced, defined as the difference between the estimated ECP (ECP_{est}) and the theoretical equilibrium potential (E_{eq}):

$$\Delta ECP = ECP_{est} - E_{eq} \quad (1)$$

The E_{eq} was determined based on the magnetite (Fe_3O_4) to hematite (Fe_2O_3) phase transition.

3. Results and Discussion

The independent CatBoost model exhibited robust predictive performance across the integrated database (R^2=0.802). However, the most significant findings emerged from the SHAP analysis.

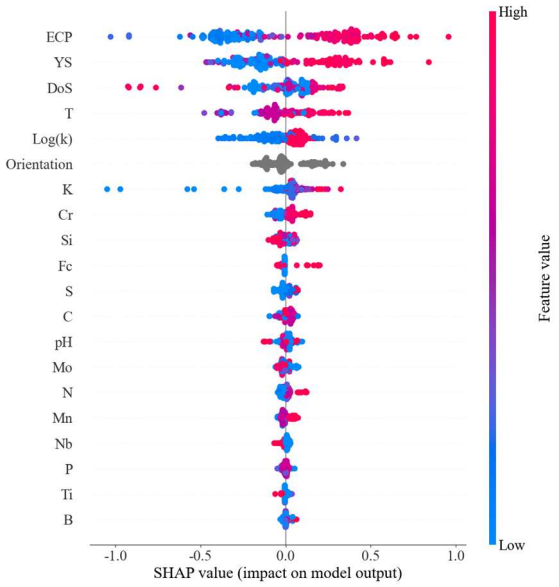
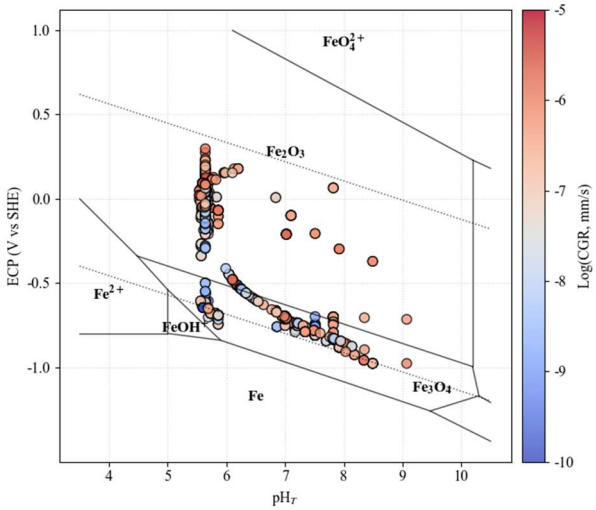
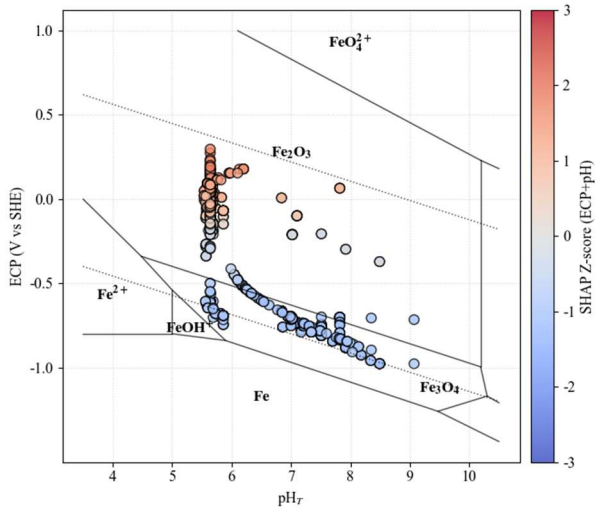


Fig. 1. SHAP feature importance plot of the CatBoost model for the combined database.

When SHAP Z-scores (Z_ϕ) were projected onto a 300°C isothermal E-pH diagram (Fig. 2) [4], the XAI model successfully captured thermodynamic behaviors without explicitly incorporating theoretical constraints. The data-driven patterns inherently assigned negative Z-scores (CGR retardation) to the stable Fe_3O_4 region and positive Z-scores (CGR acceleration) to the Fe_2O_3 region.

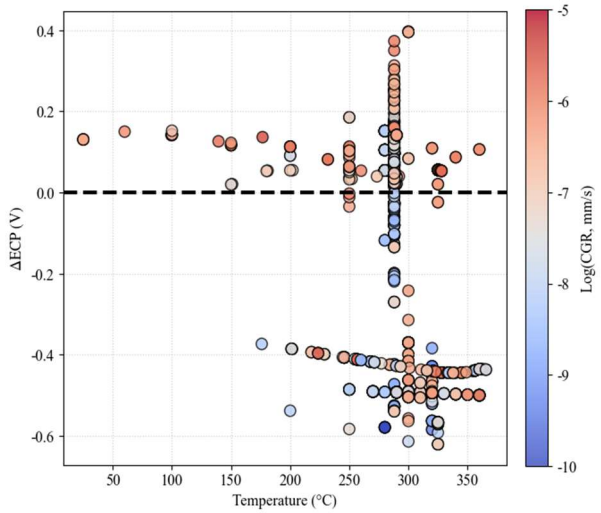


(a) Experimental CGRs

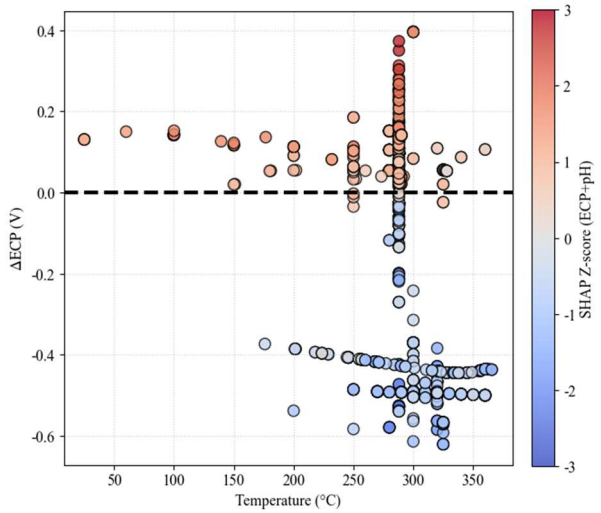


(b) SHAP Z-scores (Z_ϕ)

Fig. 2. Comparison of (a) experimental CGRs and (b) SHAP Z-scores on the Fe-H_2O E-pH diagram at 300°C.



(a) Experimental CGRs



(b) SHAP Z-scores (Z_ϕ)

Fig. 3. Comparison of (a) experimental CGRs and (b) SHAP Z-scores on *T*-ΔECP maps.

The capability of XAI to reveal thermodynamic principles was further highlighted in the dynamic ΔECP analysis (Fig. 3). While the raw experimental CGR data (Fig. 3a) showed a complex distribution influenced by mechanical and material factors, the internal decision-making criteria of the ML model (Fig. 3b) revealed a distinct boundary precisely at ΔECP = 0. The model clearly identified that regions where Fe_2O_3 is stable (ΔECP > 0) accelerate CGR, whereas Fe_3O_4 stable regions (ΔECP < 0) retard CGR. This demonstrates that XAI can successfully decouple thermodynamic effects from multifaceted empirical databases.

4. Conclusions

This study demonstrated that applying an XAI framework enables the extraction of thermodynamic patterns that are otherwise difficult to identify in complex empirical databases. By bridging machine learning outputs with E-pH diagrams and Δ ECP concepts, we confirmed that the model effectively identifies critical phase boundaries that drive material degradation. This approach highlights XAI not merely as a predictive algorithm, but as a powerful analytical tool for thermodynamic interpretation, enabling the discovery of underlying physical principles in nuclear materials research.

Acknowledgements

This work was supported by the Basic Research Support Program (2 years) of Pusan National University.

REFERENCES

[1] D.F. Falaakh, C.B. Bahn, Machine learning for modeling intergranular stress corrosion cracking of stainless steels in light water reactors with uncertainty quantification and explainability, Npj Mater Degrad 10 (2025) 5. https://doi.org/10.1038/s41529-025-00717-0.

[2] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, CatBoost: unbiased boosting with categorical features, Advances in Neural Information Processing Systems 31 (2018) 6638–6648.

[3] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems 30 (2017) 4765–4774.

[4] B. Beverskog, I. Puigdomenech, Revised pourbaix diagrams for iron at 25–300 °C, Corrosion Science 38 (1996) 2121–2135. https://doi.org/10.1016/S0010-938X(96)00067-4.