Machine learning for modeling IGSCC in austenitic stainless steels in LWRs

Dayu Fajrul Falaakh[a], Chi Bum Bahn[a*]
[a] *School of Mechanical Engineering, Pusan National University, Busan 46241, Republic of Korea*
[*] *Corresponding author: bahn@pusan.ac.kr*

Keywords : machine learning, IGSCC crack growth, austenitic stainless steels

1. Introduction

Extensive laboratory tests have been conducted to study stress corrosion cracking (SCC) in austenitic stainless steels (SSs) in light water reactor (LWR) environments, including in high temperature pure water and/or Li and B added water environments, providing abundant amount of stress corrosion crack growth rate (CGR) data. However, to date, no effort has been devoted to analyzing the compiled CGR data collected from multiple investigators.

In this work, a machine learning (ML) aided modeling approach is proposed to study the extensive data of SCC CGR in austenitic SSs. The approach includes the following steps:

1. Data preparation and preprocessing to convert the raw data into a readable form for training ML models.
2. Preliminary data analysis, i.e., correlations among input variables (i.e., features).
3. Training ML models using 10-fold cross validation scheme.
4. Interpretation of ML models using the SHAP method

2. Methods

2.1 Data preparation and preprocessing

In this work, there are 1130 experimental data on SCC CGRs in austenitic stainless steels extracted from literature [1]. The data set consists of 661 tests in boiling water reactors (BWRs) and 469 tests in pressurized water reactors (PWRs). Most tests were conducted under constant loading conditions (1041 tests), while few others were conducted under "gentle" cyclic loading conditions (89 tests). The tested materials include 17 heats of Type 304 SS (559 tests), 12 heats of Type 316 SS (429 tests), 6 heats of Type 347 SS (78 tests), and only 2 heats of Type 321 SS (64 tests). It is noted that the data set does not include tests of weld or HAZ materials.

Prior to the analysis, the data set must be preprocessed to optimize the learning process of ML models. Data preprocessing is an important phase of data-driven predictive analysis and includes several steps including data scaling, encoding, and missing value imputation.

For data scaling, the CGRs are transformed into logarithmic scale. For data encoding, categorical features are transformed into numerical features using one-hot encoding. Features with missing values are imputed using the mean value (e.g., chemical compositions) and conservative value (e.g., crack orientation relative to deformation). There are other critical features with missing values, such as pH, water conductivity, ECP, and yield strength. For pH and conductivity, their values are estimated by computing the concentrations of chemical species in the solution (see Ref.2). The ECP values are estimated using the mixed-potential theory [2]. The yield strength values are estimated using a ML model based on an extensive database [3].

2.2 CatBoost

CatBoost4 is a boosting algorithm with powerful predictive capability. To implement this algorithm, the CatBoost Python package [4] is used. To optimize the model, two key hyperparameters (i.e., the number of trees and maximum tree depth) are tuned using grid search method [5], while the default values from the ML package are assigned for other hyperparameters. The model is trained and tested in a 10-fold cross validation scheme.

2.3 Shapley additive explanation (SHAP)

The Shapley Additive explanation (SHAP) method is proposed by Lundberg and Lee [6]. The SHAP method computes the Shapley values for input features per individual prediction by assuming a linear model of feature coalition. It is assumed that a model prediction starts from a baseline value. The Shapley value for each feature increases or decreases the prediction from the baseline. To implement the SHAP method, the SHAP Python package [6] was used in this study.

3. Results

3.1 Prediction of CGR

The grid search method suggests that the optimum number of trees and tree depth for the model are 1500 and 8, respectively. Fig. 1 shows the parity plot, comparing the logarithm of actual (measured) and predicted CGRs by CatBoost. If the predictions are "error-free", the points should fall on the diagonal line in the plot. The measured mean absolute error (MAE) for the test set is 0.236 ± 0.017. This indicates that the predicted CGRs are within a factor of ~1.7 on average to the actual CGRs in linear scale. R2 for the test set is 0.836 ± 0.043, which implies satisfactory model's performance.

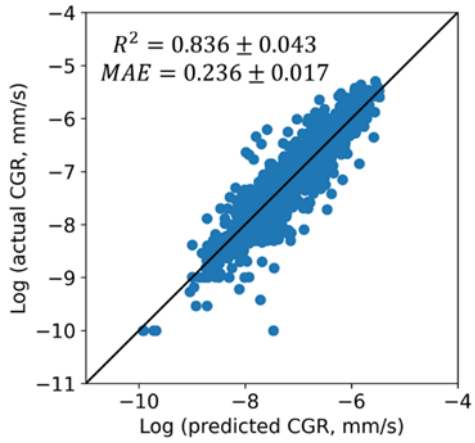

Fig. 1. Parity plots comparing the predicted and actual CGRs on the test sets in the cross-validation scheme.

3.2 Model interpretation

The ECP, K (stress intensity factor), YS_TT (yield strength at test temperature), DOS_Pa (degree of sensitization), and T (temperature) are the most important features (Fig. 2). Features such as pH_TT (pH at test temperature), crack orientation (i.e., SL and TL), and k_TT (water conductivity at test temperature) are moderately important to the predicted CGRs.

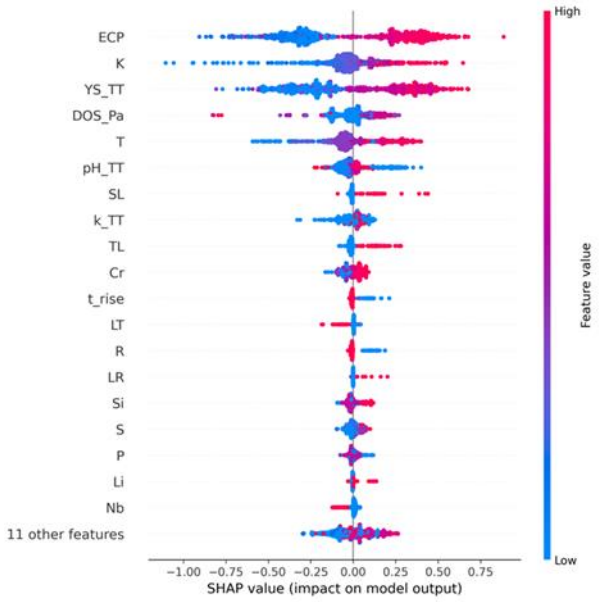

Fig. 2. SHAP values for each feature.

The ECP, K, YS_TT, and T appear to have positive correlations with their SHAP values, indicating that increasing these features accelerates the CGR. The t_rise (rise time) and R (load ration) have clear negative effects, showing that "gentle" cyclic loading may yield a SHAP value up to ~0.5 in total or accelerate the CGR up to a factor of 3 in linear scale. The DOS_Pa, pH_TT and k_TT have wide ranges of SHAP values (> 0.5 or a factor of > 3 in linear scale), but the patterns are not quite clear. The SL and TL orientations appear to have moderate effects (accelerate the CGR up to a factor of 3). Chemical compositions are also found to have weak effects on the CGR. The positive correlation of chromium content (Cr) indicates that the CGR is slightly faster in Type 304 SS than in Type 316 SS.

3. Conclusions

A workflow of ML implementation for making accurate and efficient predictions of SCC CGR in austenitic SSs is developed in this work. Based on the model interpretation using the SHAP method, the importance and influence of each feature on the model predictions are analyzed. This approach successfully provides insight into the effects of key factors on the CGR. The implementation of this workflow can be expected to accelerate the process of understanding the mechanism of this degradation.

REFERENCES

[1] Ehrnstén, U., Andresen, P.L. and Que, Z., 2023, "A review of stress corrosion cracking of austenitic stainless steels in PWR primary water", Journal of Nuclear Materials, pp.154815.
[2] Falaakh, D. F., Bahn, C. B., 2024 "Calculation of pH, conductivity, and corrosion potential of stainless steels in high temperature water chemistry conditions," Trans. of the KPVP, to be published.
[3] Materials Algorithms Project. Available online: https: //www.phase-trans.msm.cam.ac.uk/map.
[4] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A, 2018, "CatBoost: unbiased boosting with categorical features", Adv. Neural Inf. Process Syst., pp. 6638–48.
[5] Shekar BH, Dagnew G., 2019, "Grid search-based hyperparameter tuning and classification of microarray cancer data", In: 2019 2nd Int. Conf. Adv. Comput. Commun. Paradig. ICACCP, pp. 1–8.
[6] Lundberg S. M., Lee S-I., 2017, "A unified approach to interpreting model predictions", Adv. Neural Inf. Process Sys, Vol. 20, pp.4765–4774.