Machine Learning Prediction of Fracture Toughness in Hydrogen-charged Stainless Steels

Eric Hoar,* Dale Hitchcock, Timothy Krentz, Lindsay Roy

Savannah River National Laboratory, Aiken, SC 29808, eric.hoar@srnl.doe.gov

[leave space for DOI, which will be inserted by ANS]

INTRODUCTION

Austenitic stainless steels are structural materials utilized in tritium gas pressure boundaries since they are resistant to hydrogen isotope embrittlement [1-3]. However, exposure to tritium over long periods of time leads to tritium uptake which decays to result in helium ingrowth. This helium ingrowth results in further embrittlement effects which are synergistic with that from the hydrogen isotope [4]. Therefore, it is important for tritium facilities to understand the material limitations of stainless steel in this environment. The Savannah River National Laboratory (SRNL) has available a large experimental data set of austenitic stainless steels which have been exposed to tritium environments for various lengths of time. With the availability of this data set, machine learning (ML) algorithms provide an opportunity to model the embrittlement of stainless steel due to the algorithm's ability to identify patterns in data sets that are difficult and costly to identify in other manners [5]. Ultimately, the amount and quality of the available data is one defining force in the ability of a ML model to accurately predict the desired outputs. The models developed herein will illustrate the ability for the various algorithms to predict the change in fracture toughness in stainless steels due to hydrogen-isotope embrittlement.

Methodology

Over the past two and a half decades SRNL has maintained hydrogen- and tritium-charged samples with various hold times ranging from zero days to 5 years [2,6-15]. These samples correspond to 215 separate datapoints and include as-received, hydrogen-charged and tritium-charged samples. For each sample the information obtained included steel type, alloying element content, charging isotope (none, hydrogen, tritium), exposure temperature, exposure over-pressure, age time, age temperature, hydrogen content, tritium content, helium content, initial yield strength, and initial tensile strength. Figures 1 illustrates the experimental data obtained from two of the reports and show the influence of He content on the fracture toughness [9]. Table I provides a full list of the variables considered in this study and provides the approximate range of each variable. It is important to note that some variables, like the alloying element contents, are being processed over a very limited range. This can reduce the effectiveness of the machine learning model to learn the influence of the given parameter on the resultant fracture toughness.

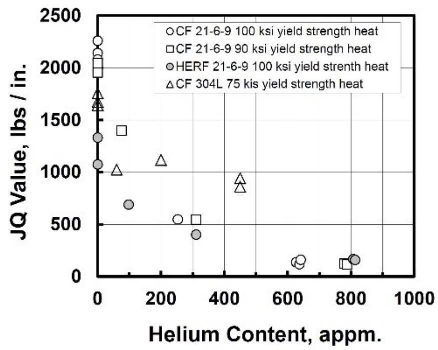

Fig. 1. Fracture Toughness vs Helium Content obtained from [9] showing some of the experimental datapoints in utilized in this study.

TABLE I. Variables of interest and their approximate range.

Variable	Value Range
Cr	0-21 wt.%
Ni	6-13 wt.%
Mn	0.5-10 wt.%
Mo	0-3 wt.%
C	0-0.05 wt.%
Si	0.3-0.7 wt.%
Cu	0-0.015 wt.%
P	0-0.03 wt.%
S	0.001-0.007 wt.%
N	0-0.3 wt.%
Co	0-0.07 wt.%
O	0-0.005 wt.%
Al	0-0.03 wt.%
YS_0	296-722.57 MPa
TS_0	510-961.13 MPa
H Content	0-6300 appm
T Content	0-5560 appm
He Content	0-1450 appm

A Random Forest (RF) machine learning algorithm was developed to predict the fracture toughness of stainless steels after hydrogen-isotope embrittlement. The RF model was developed using the scikit-learn Python library (version 1.0.2). Figure 2 below illustrates the comparison of training dataset size and average percent error.

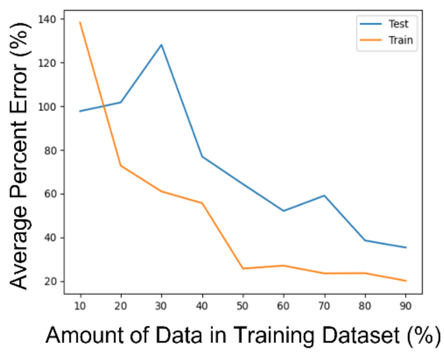

Fig. 2. Training dataset size vs Average percent error with trend lines for both testing error and training error for the RF regression model.

From Figure 2 it can be shown that for the RF model the average percent error for both the test and training datasets decrease as the size of the training dataset increases. The full embrittlement database was divided into a training dataset, a testing dataset, and a validation dataset. The training dataset comprised 80% of the full database while the testing and validation datasets each comprised 10% of the full database. The 80% size was chosen due to Figure 2 where it was identified as the point where the average percent error approached a constant value. The RF model was then optimized using the built-in function RandomizedSearchCV on the training dataset. This function utilizes a search space defined by the user to identify the optimal model hyperparameters for the desired algorithm. The optimized hyperparameters were then used as inputs to the Random Forest Regression algorithm provided by scikit-learn. The model was trained on 10 different sets of optimized hyperparameters with each set of hyperparameters being trained and tested on 10 different training and testing splits. The resultant error analysis was performed for each split and averaged across all 10 splits. After training and testing the final, best model was used to predict the fracture toughness for a validation dataset. The results of the model are discussed in the next section.

RESULTS

Once the hyperparameters were optimized the models were used to predict the fracture toughness of the training dataset. These results were used to develop plots of actual fracture toughness vs predicted fracture toughness. Figure 3 illustrates the initial results for the RF model where the solid green line identifies the optimal prediction. If the model was 100% accurate then all of the data points would fall on the green optimal line. Figure 3 shows that the RF model is capable of accurately predicting the fracture toughness of various stainless steels both with and without the inclusion of the As-Received data in the dataset. These As-Received data are all datapoints in the embrittlement database that have undergone no hydrogen-isotope charging.

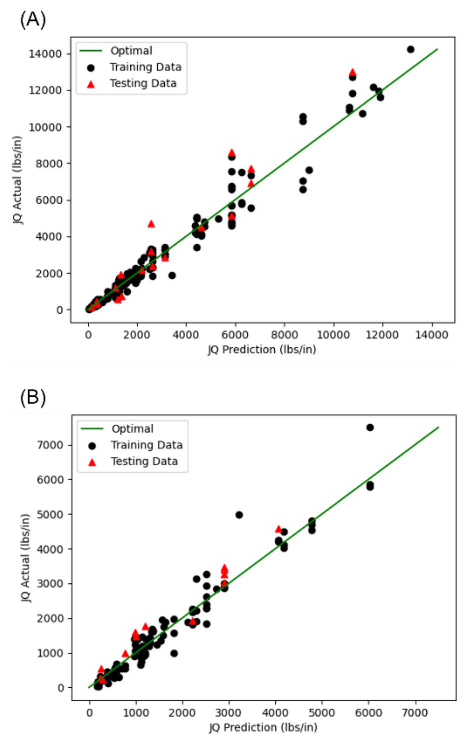

Fig. 3. Actual vs Predicted fracture toughness plots for the RF model both with (A) and without (B) the inclusion of the as-received datapoints.

In Figure 3 sections can be seen where the predicted fracture toughness is nearly the same, but the actual fracture toughness varies greatly. These sections can be identified as vertical stacks of points. When comparing these points to the experimental data in the embrittlement database it was identified that these points have the same input values but the resultant fracture toughness changes. This can be explained by the fact that fracture toughness measurements may not result in the same value for two samples due to the variability of the initial samples. In order to alleviate this issue each of the datapoints was categorized based on the values of the input variables. For each set of datapoints the fracture toughness values were overwritten using a weighted average. This means that for any two datapoints with the same input values the fracture toughness would be the same. This procedure was then performed on the embrittlement database before being used to optimize and train a secondary RF model. The results of the new model, termed the RF_W model, were used to develop new actual vs predicted plots that are illustrated in Figure 4. In Figure 4, it can be seen that the model retains its ability to accurately predict the fracture toughness, but the vertical stacks of points can no longer be seen. However, it is important to note that by performing this weighted average procedure the number of unique datapoints has been reduced.

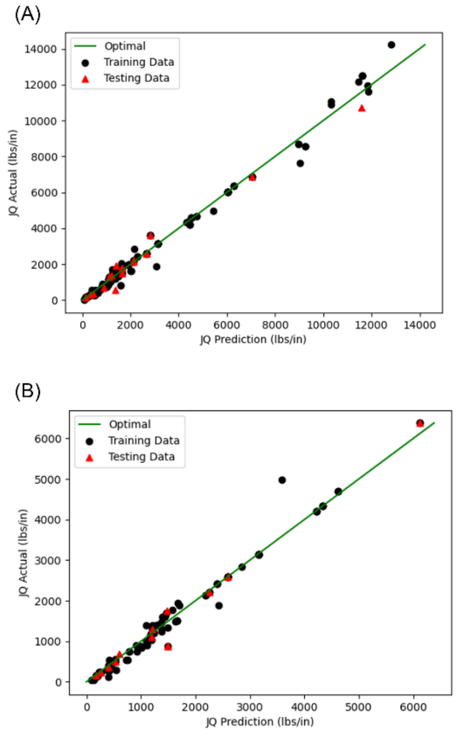

Fig. 4. Actual vs Predicted fracture toughness plots for the RF$_W$ model both with (A) and without (B) the inclusion of the as-received datapoints.

TABLE II. Comparison of the error analysis values for various metrics for the four scenarios.

Metric	RF	RF	RF$_W$	RF$_W$
As-Received?	No	Yes	No	Yes
R^2	0.9628	0.9588	0.9872	0.9953
MAE Train (lbs/in)	190.696	368.204	77.017	103.144
MAPE Train (%)	16.820	18.181	8.816	10.362
MAE Test (lbs/in)	266.209	527.246	216.516	157.207
MAPE Test (%)	28.154	44.765	20.842	16.414
MAE Validate (lbs/in)	288.469	382.467	116.332	226.221
MAPE Validate (%)	32.213	23.729	19.690	20.312

The error analysis performed here is provided in Table II below. From Table II it can be seen that both models perform very well with high R^2 values and low MAPE and MAE scores. As expected, the training MAE and MAPE scores are also lower than the test scores for all cases. However, the MAE and MAPE scores for the validation set were identified as being higher than the testing scores for some of the models meaning overfitting may be an issue in the model Furthermore, it is important to note that the error analysis results for all scenarios indicate that the model may be overfitting the training dataset. This overfitting could be due to the limited number of training data utilized in this study. Therefore, a utilizing a larger dataset for future results is paramount to be able to identify the true ability of the RF model for predicting fracture toughness. Meanwhile, Figures 5 and 6 provide four plots in which the RF$_W$ model was used to predict the fracture toughness for samples obtained from a singular reference. This was done to provide a comparison of the predicted value with the trends seen in the experimental data. Figure 5 compares the predicted fracture toughness vs H content trend line to experimental results for the F97, H94, F9 and 98 samples obtained from [6]. Furthermore, Figure 6 compares the predicted fracture toughness vs He content trend line to experimental results for the same samples obtained from [6]. The red points and stars correspond to the model's fracture toughness for a fictional H-charged sample and T-charged sample at the same charging (i.e. same appm). This was done to identify if the model is capable of discerning a difference between the H-charged samples and the T-charged samples (without He content).

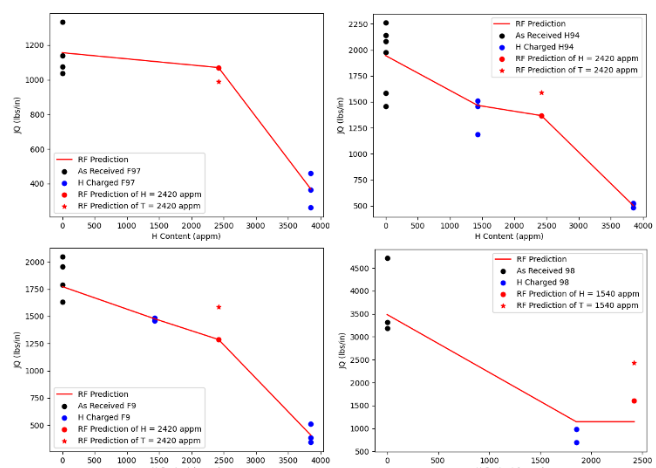

Fig. 5. Fracture toughness vs H-content plots comparing the prediction obtained from the RF$_W$ model to the experimental data obtained from [6].

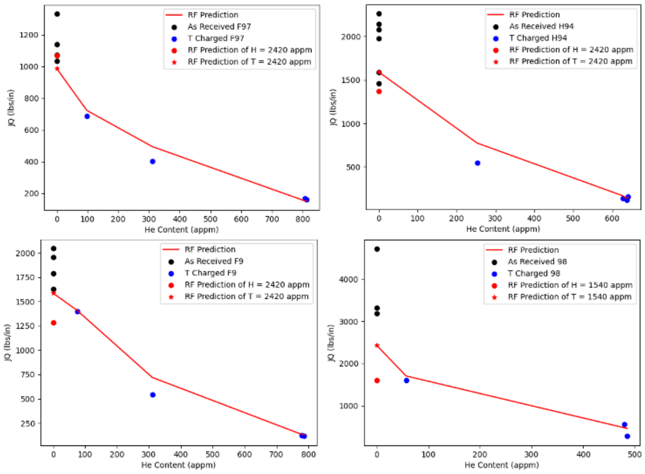

Fig. 6. Fracture toughness vs He-content plots comparing the prediction obtained from the RF_W model to the experimental data obtained from [7].

From Figures 5 and 6 it can be seen that the RF_W model tends to overestimate the fracture toughness values for the H-content plots while this overestimation is not seen in the He-content plots. However, in all plots the model is able to accurately predict the fracture toughness for most of the experimental datapoints. One possible explanation for this is that the RF_W model is currently overfitting the data. This can occur due to the limited size of the current embrittlement dataset and can be rectified by increasing the data available.

In conclusion, a Random Forest regression algorithm was developed to predict the resultant fracture toughness of various hydrogen-isotope embrittled stainless steels. Decades worth of experimental data on tritium charged stainless steels was obtained from the Savannah River National Laboratory. The data was combined into a singular repository of approximately 215 data points for which 80% were used as a training set, 10% were used as a testing set and 10% were used as a validation set. Two preprocess techniques were utilized to optimize the database for use in the RF model. The error analysis performed herein illustrated that the RF model was capable of predicting fracture toughness. However, the error analysis also illustrated that the data obtained herein was insufficient to satisfy the null hypothesis and as such more data is required it also. Ultimately, the results presented herein are preliminary results obtained from the implementation of a Random Forest model to a limited embrittlement database and future work will be required to identify the full capabilities of the model and to overcome the difficulties with overfitting that have been seen.

NOMENCLATURE

SRNL = Savannah River National Laboratory
T = Tritium
ML = Machine Learning
RF = Random Forest
YS_0 = Initial Yield Strength
TS_0 = Initial Tensile Strength
RF_W = Random Forest using Weighted Data
MAE = Mean Absolute Error
MAPE = Mean Absolute Percent Error

REFERENCES

1. G. Caskey Jr, "Tritium-helium effects in metals," Fusion Technology, **8**: 2293-2298 (1985).
2. M. J. Morgan, M. H. Tosten and S. L. West, WSRC-STI-2006-00056, Savannah River National Laboratory.
3. M. J. Morgan, D. A. Hitchcock, T. M. Krentz and S. L. West, "Tritium Aging Effects on Fracture Toughness of Stainless Steel Weldments," Fusion Science and Technology, **76(3)**: 209-214 (2020).
4. T. M. Krentz, J. A. Ronevich, D. K. Balch and C. San Marchi, "Tritium embrittlement of austenitic stainless-steel tubing at low helium contents," Fusion Engineering and Design **168** (2021).
5. Q. Zhou, S. Lu, Y. Wu and J. Wang, "Property-Oriented Material Design Based on a Data-Driven Machine Learning Technique," Journal of Physical Chemistry Letters **11**: 3920-3927 (2021).
6. M. J. Morgan, S. L. West and G. K. Chapman, WSRC-TR-2007-00244, Savannah River National Laboratory.
7. M. J. Morgan and G. K. Chapman, SRNL-TR-2009-00468, Savannah River National Laboratory.
8. M. J. Morgan and G. K. Chapman, SRNL-STI-2011-00726, Savannah River National Laboratory.
9. M. J. Morgan and G. K. Chapman, SRNL-TR-2014-00014, Savannah River National Laboratory.
10. M. J. Morgan, SRNL-STI-2015-00103, Savannah River National Laboratory.
11. M. J. Morgan, SRNL-STI-2016-00060, Savannah River National Laboratory.
12. M. J. Morgan, SRNL-STI-2017-00052, Savannah River National Laboratory.
13. M. J. Morgan, SRNL-STI-2017-00416, Savannah River National Laboratory.
14. M. J. Morgan, D. Hitchcock, T. Krentz, J. McNamara and A. Duncan, SRNL-STI-2018-00036, Savannah River National Laboratory.
15. T. M. Krentz and D. A. Hitchcock, SRNL-STI-2020-00002, Savannah River National Laboratory.