crystals

MDPI

Article

Machine Learning-Assisted Prediction of Stress Corrosion Crack Growth Rate in Stainless Steel

Peng Wang [1,2,*], Huanchun Wu [1,2], Xiangbing Liu [1,2] and Chaoliang Xu [1,2]

[1] Material Engineering Technology Center, Suzhou Nuclear Power Research Institute, Suzhou 215004, China; wuhuanchun1@163.com (H.W.); liuxbing@cgnpc.com.cn (X.L.); xuchaoliang@cgnpc.com.cn (C.X.)
[2] National Engineering Research Center for Nuclear Power Plant Safety & Reliability, Suzhou 215004, China
* Correspondence: wang_peng@cgnpc.com.cn

Abstract: Stainless-steel is extensively utilized in the key structural components of the main equipment in the nuclear island of pressurized water reactor nuclear power plants. The operational experience of nuclear power plants demonstrates that stress corrosion is one of the significant factors influencing the long-term safe operation of stainless steel in the high-temperature water of pressurized water reactor nuclear power plants. This study is based on the stress corrosion crack growth rate data of 316SS and 304SS stainless steel in the simulated primary water environment of pressurized water reactor nuclear power plants. Data mining and modeling were conducted using multiple machine learning algorithms, including Random Forest (RF), eXtreme Gradient Boosting (XGBoost), Support Vector Regression (SVR), and Gaussian Process Regression (GPR), and the Sharpley Additive explanation (SHAP) method was employed to analyze the interpretability of the model. The results indicate that the stress corrosion crack growth rate prediction model based on XGBoost outperforms other models in all assessment indicators. Compared with empirical equations, XGBoost exhibits high flexibility and excellent data-driven learning capabilities. In the test set, 90% of the prediction errors are within the range of experimental values, with the maximum error multiple being 2.5, which significantly improves the prediction accuracy. Moreover, the distribution of SHAP values is consistent with the theoretical study of the stress corrosion behavior of stainless-steel, effectively reflecting the impact of cold working, temperature, and stress intensity factor on the stress corrosion crack growth rate, thereby proving the reliability of the model's prediction results. The achievements of this study hold significant reference value and application prospects for the prediction of the stress corrosion behavior of stainless-steel in a high-temperature and high-pressure water environment of pressurized water reactor nuclear power plants.

Keywords: stress corrosion cracking (SCC); 316 and 304 stainless steel; pressurized water reactor (PWR); machine learning

check for updates

Citation: Wang, P.; Wu, H.; Liu, X.; Xu, C. Machine Learning-Assisted Prediction of Stress Corrosion Crack Growth Rate in Stainless Steel. *Crystals* **2024**, *14*, 846. https://doi.org/10.3390/cryst14100846

Academic Editor: José L. García

Received: 1 August 2024
Revised: 22 August 2024
Accepted: 27 August 2024
Published: 27 September 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The safe and efficient operation of nuclear power plants is inseparable from the reliability of the service of nuclear power materials. Stainless steel is widely used in the main equipment, pipelines, and welds of the nuclear island of pressurized water reactors, due to its outstanding mechanical properties and corrosion resistance [1,2]. Operating experience indicates that stress corrosion cracking (SCC) is one of the important factors affecting the long-term safe operation of stainless steel in the high-temperature water of pressurized water reactors (PWR), and the failure accidents of key components caused thereby seriously threaten the safe operation of nuclear power plants [3–6]. Therefore, establishing a prediction model for the SCC behavior of stainless steel in a simulated high-temperature and high-pressure water environment of nuclear power plants is of great significance for the safety evaluation of the long-term service of nuclear power plants.

The SCC behavior of nuclear power materials in high-temperature water is the result of the coupling effect of materials, stress, and the environment [7,8]. Currently, SCC crack

propagation prediction models can be divided into two categories. One is theoretical prediction models, proposed based on the SCC mechanism, such as the slip oxidation model [9], FRI model [10], coupled environment fracture model (CEFM) [11], internal oxidation model [12], etc., which have good interpretability of SCC phenomena; the other is empirical prediction models established based on a large amount of laboratory data and engineering data, such as the PMScott model [13], MRP-55 model [14], etc., which have a strong application basis. Due to the fact that there are more than 20 influencing factors for SCC behavior in high-temperature and high-pressure water environments [12–17], and there are multi-factor coupling effects among various parameters, the existing parametric models either cannot accurately express the correlation between each influencing factor, or are highly complex with numerous parameters, making them difficult for engineering applications. Therefore, how to establish an SCC prediction model with wide coverage of influencing factors, wide applicability, and high accuracy is the current challenge.

Machine learning has a powerful ability to discover and process information. It directly studies the data through algorithms to obtain the dependencies between the data, thereby making predictions and judgments on unknown data. Because machine learning is data-driven and has no fixed model structure requirements, it is more applicable to handling complex problems compared to theoretical or empirical (semi-empirical) methods [18–22]. Therefore, in recent years, data-driven machine learning has gradually been more widely used in the prediction of material service behavior. Churyumov et al. [18] constructed an artificial neural network (ANN) model with high accuracy to describe the high Mn steel deformation behavior in dependence on the concentration of the alloying elements, the deformation temperature, the strain rate, and the strain. Li et al. [19] established four different prediction models of machine learning algorithms, including Support Vector Regression based on radial basis kernel function (SVM-RBF), ridge regression (RR), Random Forest (RF), and back propagation neural network (BPNN), to predict the formation energy, thermodynamic stability, crystal volume, and oxygen vacancy formation energy of perovskite materials. Mei et al. [21] developed a Knowledge-Based Random Forest (KBRF) model to predict the SCC growth rate of the nicked-based 690 alloy through combining RF with a domain knowledge-based MRP-386 parameterized model. Rao et al. [22] combined the Gaussian mixture models with the density-based spatial clustering of applications with a noise clustering algorithm to process the orientation data acquired through electron backscatter diffraction, establishing a machine learning method for the segmentation and characterization of the micro-texture regions of near-α titanium alloys.

This study aimed at the SCC problem of stainless-steel materials under the service environment conditions of the PWR primary water. Four machine learning algorithms were utilized to establish the prediction model of the SCC crack growth rate of stainless steel, and the process is shown in Figure 1. First, the dataset of the SCC crack growth rate of stainless steel was randomly divided into two parts. Among them, 70% of the data was used as the training set for the training of the machine learning model, and the remaining 30% of the data was used as the test set to evaluate the performance of the model. Next, the method combining the eXtreme Gradient Boosting (XGBoost) feature importance ranking and the Pearson correlation coefficient was adopted for feature selection. Four machine learning algorithms, namely RF, XGBoost, Support Vector Regression (SVR), and Gaussian Process Regression (GPR), were used for data mining to establish the prediction model of the SCC crack growth rate of stainless steel, and the prediction accuracy of the models was compared and evaluated. Finally, the Sharpley Additive explanation (SHAP) method based on game theory was used to explain the best model and explore the influence of each feature on the model's prediction results. The achievements of this study have important reference value and application prospects for the prediction of the stress corrosion behavior of stainless steel in the high-temperature and high-pressure water environment of PWR.

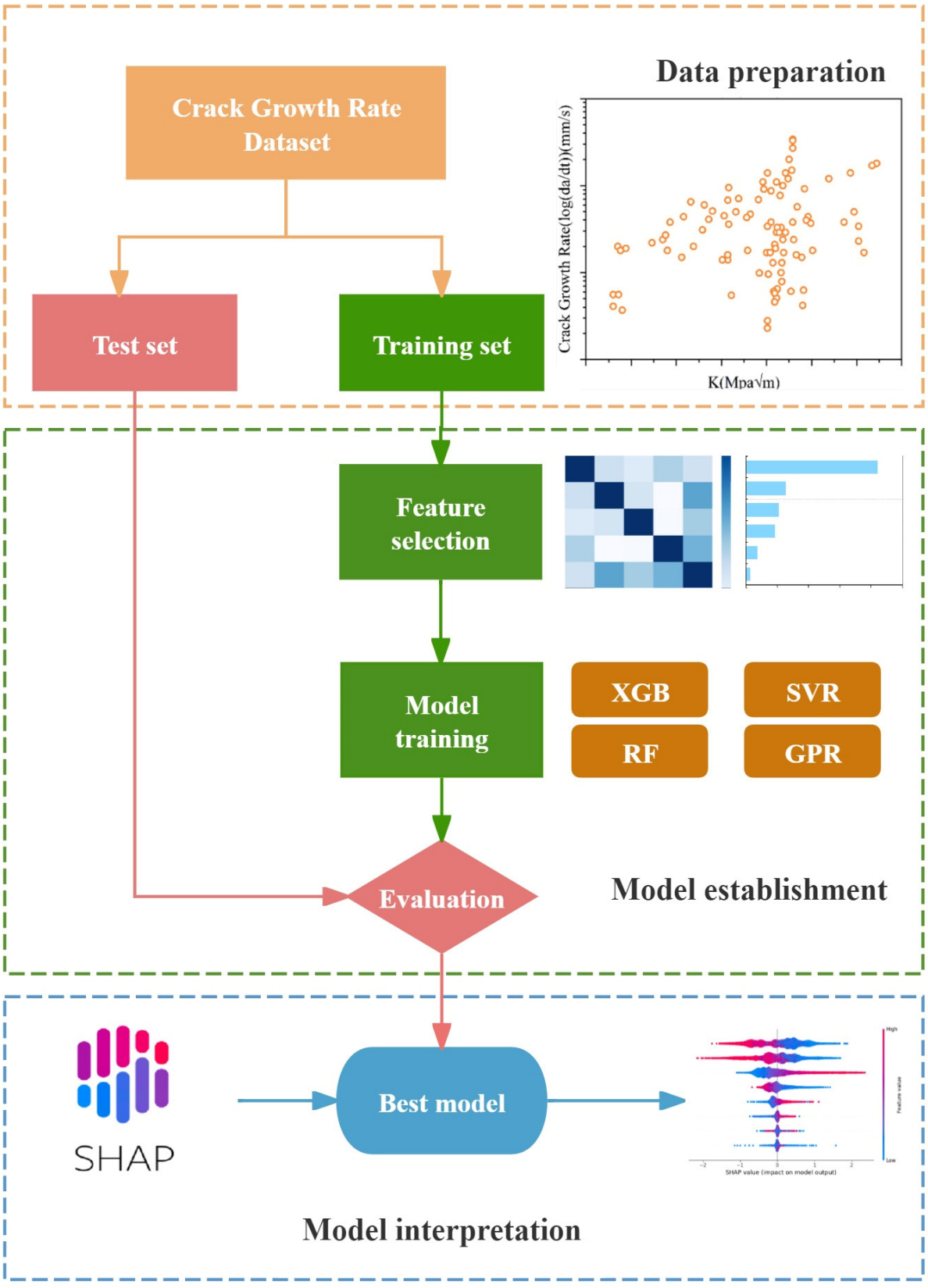


Figure 1. Flow chart for predicting stress corrosion crack growth rate based on machine learning.

2. Methods

2.1. Dataset

During the manufacturing process of nuclear power equipment and structural components, to enhance the strength of the materials they are typically subjected to cold working (CW) treatment. The operational experience from nuclear power plants has indicated that materials with CW have a higher tendency for stress corrosion cracking (SCC) [23,24]. Investigating the SCC performance of pressure boundary materials after CW in the primary coolant environment is of significant importance for the safe operation and life cycle management of nuclear power stations. Terachi et al. [25] have systematically quantified the SCC crack growth rate of CW stainless steel in a simulated PWR primary water environment. The data used in this study originate from the research conducted by [25], comprising a total of 99 data points, which are shown in Table A1. The elemental composition of the stainless steel used in the experiments is shown in Table 1, and the characteristics of the data are presented in Table 2. The materials were solution-treated at 1060–1080 °C for 10 min then water quenched. To examine the influence of CW, the materials were cold-rolled in one dimension to produce 5%, 10%, 15%, and 20% reductions in thickness at room temperature. The SCC test uses 0.5T compact tension specimens and is carried out in simulated PWR primary water that contained 500 ppm boron as boric acid, 2 ppm lithium as LiOH and 30 cm^3-STP/kg-H_2O of dissolved hydrogen. The crack growth rate is measured by the direct current potential drop method described in ASTM standard E647-11. Prior to the crack growth rate measurement, each specimen was pre-cracked in air by fatigue stress by applying a symmetrical triangle wave equal to or less than 10 Hz.

Table 1. Compositions of alloys (mass%).

Element	C	Si	Mn	P	S	Ni	Cr	Mo
316SS	0.047	0.45	1.42	0.024	0.001	11	16.45	2.07
304SS	0.04	0.31	1.59	0.031	0.001	9.21	18.34	0.37

Table 2. Characteristic parameters of the stainless-steel SCC crack growth data set.

No.	Characteristic Parameters	Value
1	Cold Work	5–20%
2	Yield Strength	256–587 MPa
3	Vickers Hardness	184–270 HV
4	K	13–42.2 Mpa$\sqrt{m}$
5	Temperature	250–360 °C
6	Crack Growth Rate	2.3×10^{-9}–3.4×10^{-7} mm/s

2.2. Machine Learning Algorithm

2.2.1. RF

RF is an ensemble learning algorithm based on decision trees. It uses the random repeated sampling technique to randomly and repeatedly sample the data set Hi with replacement to obtain N subsets θ_k each containing m samples. For each subset, a decision tree $h(x_i, \theta_k)$ is established. All the decision trees are integrated to form RF. The arithmetic mean of the prediction results of the N decision trees is calculated as shown in Equation (1), to obtain the RF prediction output [26]. The constructed model can reduce the correlation between decision trees, effectively handle high-dimensional data, has a high tolerance for outliers and noisy data, and has good generalization ability and accuracy.

$$a_{RF}^{i} = \frac{1}{N}\sum_{k=1}^{N} h(x_i, \theta_k) \quad (1)$$

2.2.2. XGBoost

XGBoost is an ensemble learning algorithm based on the theory of gradient boosting. Its principle is to integrate a large number of weak classifiers together by using the idea of iterative operation to convert them into strong classifiers, and determine the weight of each classifier by calculating residuals. Shrinkage and regularization techniques are used to handle sparsity and avoid overfitting [27,28]. After performing the corresponding t iterations on the XGBoost model, the expression is as follows:

$$\widehat{y}_i^{(t)} = \widehat{y}_i^{(t-1)} + f_t(x_i) \tag{2}$$

At this point, the objective function of XGBoost is

$$\lambda^{(t)} = L\left(y_i, \widehat{y}^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{3}$$

In the formula, x_i is the feature, y_i is the true value, $\widehat{y}_i^{(t)}$ is the predicted value, $L\left(y_i, \widehat{y}^{(t-1)} + f_t(x_i)\right)$ is the loss function, $\Omega(f_t)$ is the regularization term to avoid overfitting, and f_t is the t-th regression tree.

To prevent overfitting, the error function is expanded using the second-order Taylor formula at point $\widehat{y}_i^{(t-1)}$. After integration, the final objective function is obtained as

$$\widehat{\lambda}^{(t)} = \sum_{i=1}^{n}\left[g_i f_t(x_i) + \frac{1}{2}h_i f_i^2(x_i)\right] + \Omega(f_t) \tag{4}$$

In the formula, g_i and h_i are the first-order and second-order derivatives obtained from $\widehat{y}_i^{(t-1)}$.

2.2.3. SVR

Support Vector Machine (SVM) is an important method in data mining. It is based on the Vapnik–Chervonenkis dimension theory in statistical learning theory and the principle of structural risk minimization. Its essence is to solve the convex quadratic programming problem. SVR is a method for regression problems in SVM. SVR can obtain the global optimal solution, avoiding the problem of local extremes, and has strong generalization ability and good performance in application scenarios such as small samples and high-dimensional spaces [29,30]. Suppose the linear regression function established in the high-dimensional feature space is:

$$f(x) = \omega^T \phi(x) + b \tag{5}$$

where, b represents the bias, ω^T is the weight vector, and $\phi(x)$ is the nonlinear mapping function. Introducing the non-relaxation factor α_i^* and the Lagrange factor α_i, $\omega^T = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(x_i)$ can be obtained. After the corresponding changes, the regression function is

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{6}$$

where, $K(x_i, x)$ is the kernel function. Common kernel functions include polynomial kernel functions, radial basis kernel functions, and Sigmoid kernel functions.

2.2.4. GPR

GPR can effectively handle nonlinear multi-dimensional parameter regression problems with small sample data. The essence of this algorithm is to model nonlinear systems

based on multivariate Gaussian processes and Bayesian inference [31]. Assume that a space with any finite number of samples is $D = \{(X, y_i)|i = 1, 2, \cdots, n\} = \{X, Y_D\}$, where $X = [x_1 x_2 \cdots x_n]^T$ is an n-dimensional input matrix, $x_i \in R^k$ is a k-dimensional input vector, $y_i \in R$ is the corresponding output scalar, and Y_D is the output vector. The regression process will develop the nonlinear mapping relationship between the input X and the input Y_D based on the sample set, and this regression model is expressed as:

$$y = f(x^*) + \varepsilon \tag{7}$$

where, y is the observed value considering noise, x^* is the input vector, $f(x^*)$ is the predicted value without considering noise, ε is the noise, and $\varepsilon \sim N(0, \sigma_n^2)$. A noise of 0 indicates that Gaussian Process Regression completely interpolates the original sample. Therefore, σ_n^2 can prevent overfitting in regression.

The multivariate Gaussian process can be defined as that any input variable x, $x^\theta \in R^d$ satisfies the Gaussian process Bayesian prior distribution:

$$f(x) \sim GP\left(m(x), k\left(x, x^\theta\right)\right) \tag{8}$$

where, $m(x)$ is the mean function and $k\left(x, x^\theta\right)$ is the kernel function. In practical use, $m(x)$ is generally set to 0 to simplify the solution of the posterior probability distribution.

2.3. Feature Selection

Rich-featured datasets are beneficial for mining information from data and improving the predictive accuracy of models. However, when there are redundant features in the dataset, it will increase the complexity of model analysis and reduce the model's generalization performance. The XGBoost algorithm can calculate the optimal feature variables for the decision tree split based on the gain of the structure score, and rank the importance of the feature variables. Since this ranking result only reflects the degree of influence of a single feature variable on the predictive outcome, it cannot eliminate redundant features with significant linear relationships; Pearson's correlation coefficient is a statistical method that accurately evaluates the degree of linear correlation between two continuous variables [32]. The Pearson correlation coefficient p is defined as:

$$p = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{9}$$

where, x_i and y_i are feature variables, and $\overline{x}$ and $\overline{y}$ are the average values of n data points, respectively. The value of p is in the range $[-1, 1]$, with a positive value indicating a positive correlation and a negative value indicating a negative correlation between the two variables. The magnitude of the absolute value indicates the degree of correlation between the variables. This study uses a method that combines XGBoost importance ranking with the Pearson correlation coefficient to select feature variables.

2.4. Evaluation Method

In order to effectively evaluate the prediction accuracy of the model, the model evaluation indicators adopted are Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Median Absolute Error (MAB), and Coefficient of Determination (R^2). The smaller the MAE, MAB, and MAPE, and the larger the R^2 and the closer it is to 1, the better the prediction effect of the model.

$$\text{MAPE} = \frac{1}{n}\sum_{i}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \times 100\% \tag{10}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i}^{n}\left(y_i - \widehat{y}_i\right)^2 \tag{11}$$

$$\mathrm{MAB} = Median\left(\left|y_i - \widehat{y}_i\right|\right) \tag{12}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum_{i}^{n}\left(y_i - \overline{y}\right)^2} \tag{13}$$

where $\widehat{y}_i$ is the predicted value, y_i is the true value, and $\overline{y}_i$ is the average value of y_i.

3. Results and Analysis

3.1. Feature Selection Result

Figure 2 shows the Pearson correlation heatmap of the stainless-steel crack propagation data. The matrix values corresponding to the features are the Pearson coefficients. The darker the color of the matrix, the stronger the correlation. It can be seen that the Pearson coefficient values between all input vectors and the crack propagation rate are greater than zero, ranging from 0.12 to 0.5, indicating that they have a certain promoting effect on the growth of the crack propagation rate. The linear correlation is below the medium level, and there may be a complex nonlinear relationship. The Pearson coefficient values among CW, yield strength, and hardness are above 0.9, indicating an extremely strong linear correlation. Figure 3 shows the results of feature importance scores. The feature importance from high to low are the stress intensity factor, temperature, yield strength, CW, hardness, and carbon content. Among them, the feature importance score of yield strength is 209 points, which is greater than 184 points of the CW and 71 points of hardness. In addition, the feature importance score of carbon content is only 25 points. Therefore, when building a machine learning model, the three feature variables of CW, hardness, and carbon content are selected to be eliminated, and the final feature set is determined to be stress intensity factor, temperature, and yield strength. Based on the above analysis results, these feature parameters all have different degrees of influence on the stress corrosion crack propagation rate, which is consistent with the previous experimental research results [33–39]. For example, the increase in the stress intensity factor leads to an increase in the strain rate at the crack tip, and the crack propagation rate increases accordingly. Since the ion diffusion rate, interface reaction rate, and electrochemical potential involved in the crack propagation process are all functions of temperature, the influence of temperature T on the crack propagation rate is extremely complex. An increase in yield strength will limit the size of the plastic zone, increase the strain gradient at the crack tip, and promote stress corrosion cracking, etc.

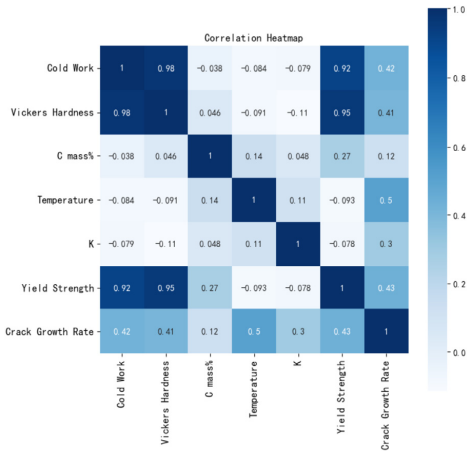


Figure 2. Pearson correlation heatmap.

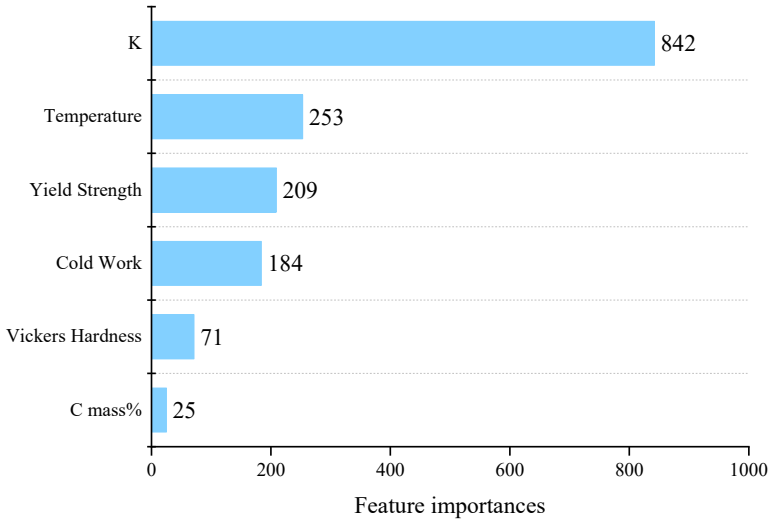


Figure 3. XGBoost feature importance ranking results.

3.2. The Result of Machine Learning Modeling

GPR, SVR, RF, and XGBoost were used to train the training set data of the SCC crack propagation rate. Grid search was used to optimize the hyperparameters of the machine learning algorithms to improve the model accuracy. The optimization results of the hyperparameters of the four machine learning algorithms are shown in Table 3.

Table 3. Optimization results of hyperparameters of machine learning algorithms.

Model	Hyperparameter
GPR	kernel = DotProduct + RBF
SVR	kernel = "RBF", C = 40
RF	max_depth = 10, n_estimators = 50, max_features = 4
XGBoost	learning_rate = 0.085, max_depth = 9, min_child_weight = 3, n_estimators = 60

The prediction performance of each model was evaluated using the test set. The evaluation indicators of each model were plotted as a radar chart, as shown in Figure 4. The closer the quadrilateral formed by each model in the radar chart is to the lower half of the coordinate axis center, the better the performance of the model. The predicted values of the model and the experimental values of the crack propagation rate were plotted as a scatter plot, as shown in Figure 5. The closer the data points are to the y = x diagonal line in the figure, the more consistent the predicted values and the experimental values are, thereby proving the better the fitting effect of the model. It can be seen that most of the predicted values of GPR are distributed in scatter band 2x, and the degree of dispersion is relatively large. GPR is based on the assumption that the objective function obeys Gaussian distribution. Stress corrosion crack growth is affected by many factors, such as material characteristics, environmental conditions, stress levels, etc. The complexity of these factors makes the distribution of crack growth data often uncertain and does not satisfy Gaussian distribution, which leads to the deterioration of the learning performance of GPR. Compared with GPR, all evaluation indexes of SVR are slightly improved. For the nonlinear prediction problem of crack growth rate, SVR uses kernel function to map the data to high-dimensional space, so as to perform linear regression in high-dimensional space. By minimizing the model complexity and error through the objective function, it helps to improve the generalization ability of the model. It can be clearly seen from Figure 4 that the model evaluation indicators of XGBoost and RF are significantly better than those of SVR and GPR. The R^2 is above 0.75, and the MSE, MRE, and MAB have decreased significantly. The data points are closer to the y = x diagonal line, indicating that the ensemble learning algorithm based on decision trees has significant advantages

in learning SCC crack propagation rate data. Compared with SVR and GPR algorithms based on a single model, it is more robust and has a better fitting effect for complex features. Among them, XGBoost adds regularization terms and gradient boosting strategies to the cost function, which can simplify the model during the training process, enhance the generalization ability of the model, and effectively prevent overfitting. Therefore, compared with RF, XGBoost has a 5% increase in R^2, a 14% reduction in MSE, a 11% reduction in MRE, and a 37% reduction in MAB. For the prediction problem of the stainless-steel stress corrosion crack propagation rate, XGBoost has higher prediction accuracy and fitting degree, and the model reliability is stronger.

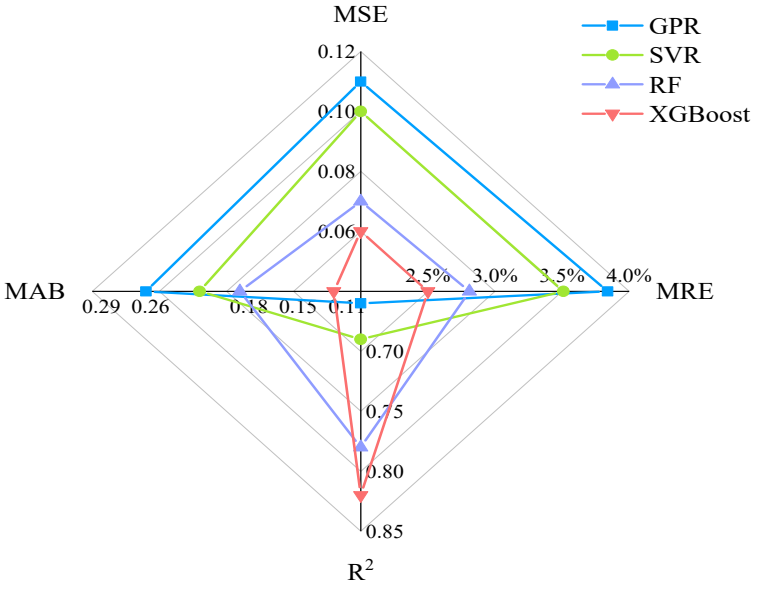


Figure 4. Radar chart.

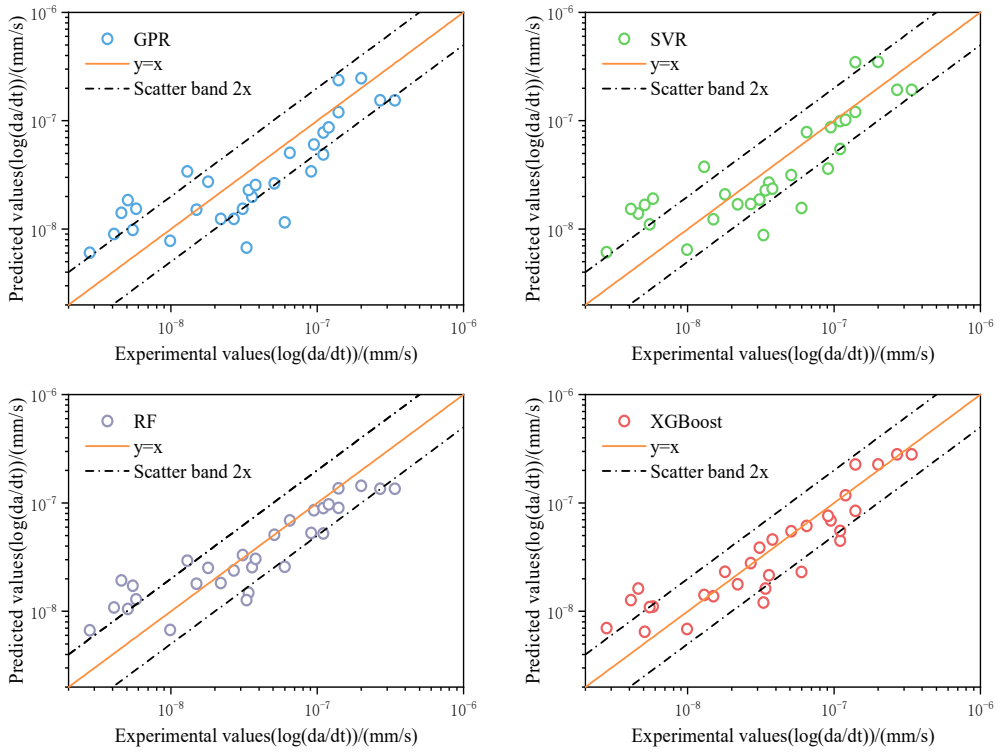


Figure 5. Comparison of model prediction results of the test set.

Terachi et al. [25] proposed an empirical equation for crack propagation rate prediction applicable in engineering based on the experimental results, as shown in Equation (14). This equation can well encompass three important engineering parameters: temperature, yield strength, and stress intensity factor. Through comparison with the accuracy of this model, the engineering applicability of the crack propagation model established based on XGBoost is further analyzed in the study.

$$\frac{da}{dt} = 9.5 \times 10^{-10} \times \sigma_y^{2.9} \times K_a^{1.0} \times \exp^{\frac{-84 \times 10^3}{RT}} \quad (14)$$

where da/dt is crack growth rate (mm/s), σ_y is yield strength at test temperature, K_a is apparent stress intensity factor, R is gas constant (8.314 kJ/mol·K), and T is absolute test temperature.

Figure 6 presents the prediction results of XGBoost and the empirical equation on the test set of stainless-steel stress corrosion crack propagation. It can be observed that XGBoost demonstrates better convergence compared to the empirical equation. Figure 7 displays the distribution frequency of the ratio between the prediction error and the experimental data. Among them, 77% of the prediction errors of the empirical equation fall within the range of the experimental value, with the maximum error ratio being 3.5 and the average error ratio being 0.75. Whereas, 90% of the prediction errors of XGBoost are within the range of the experimental value, with the maximum error ratio being 2.5 and the average error ratio being 0.5. The prediction effect of XGBoost is significantly superior to that of the empirical equation. In contrast to the empirical equation, XGBoost, based on the data-driven learning strategy, establishes a non-parametric prediction model by exploring the relationship between SCC characteristic parameters and the crack propagation rate. It can continuously adjust and optimize the model parameters and structure from a large amount of data, possessing high flexibility and excellent learning ability. However, parametric models typically require presetting the model structure. For complex nonlinear problems, it is challenging to determine the accurate expression, thereby restricting the model accuracy.

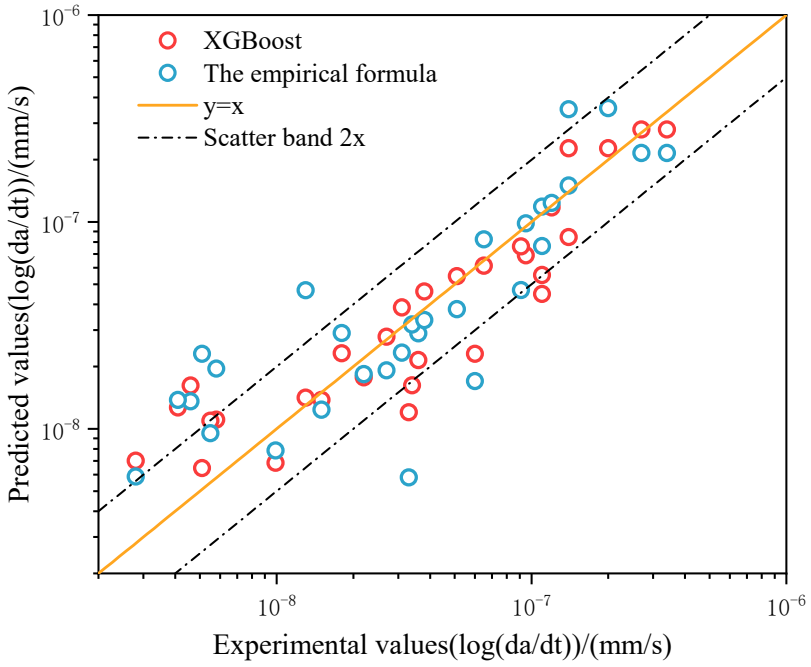


Figure 6. Prediction results of XGBoost and empirical equation on the test set.

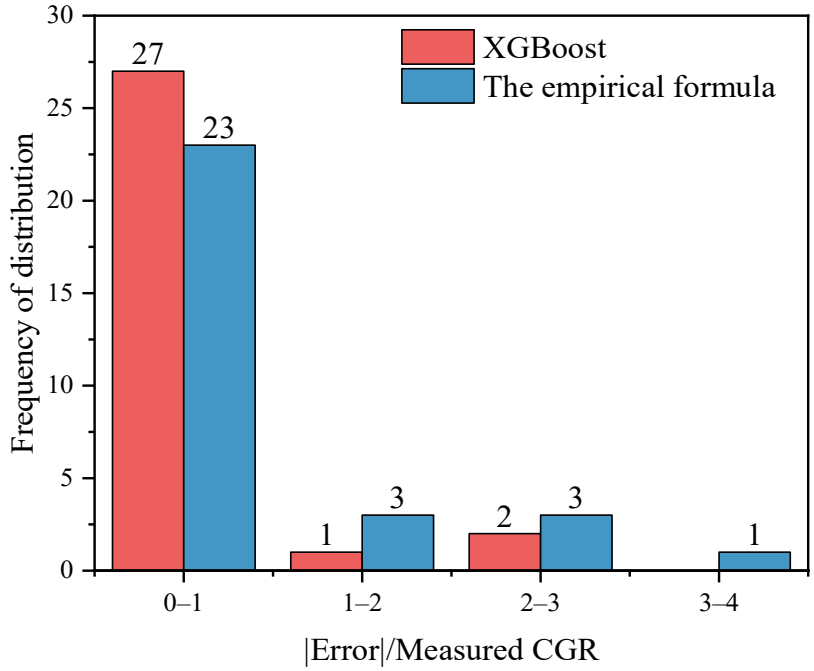


Figure 7. Distribution frequency of the ratio between prediction error and experimental data.

3.3. Model Interpretations

Although machine learning models have demonstrated excellent performance in various complex tasks, they usually operate in the form of a "black box", and users cannot understand the decision-making process inside the machine learning model, as well as the importance of feature parameters to the decision results and the interaction relationship between different feature parameters. SHAP is a global interpretation method that explains the results of machine learning models by evaluating the contribution of each input feature to the model output [40–42]. Differently from the existing feature importance attributes in machine learning, SHAP can identify whether the contribution of each input feature is positive or negative, which helps to understand the reasons why the model makes specific predictions, thereby better explaining and optimizing the model, and enhancing the credibility of the model and the acceptance of the model by users. The SHAP value obeys the following equation:

$$y_i = y_m + f(x_{i1}) + f(x_{i2}) + \cdots + f(x_{ik}) \tag{15}$$

In Equation (15), x_i represents the i-th sample, x_{ij} represents the j-th feature of x_i, y_i represents the predicted value of the model, y_m represents the baseline of the entire model (usually the mean value of the target variable of the total sample), $f(x_{ij})$ is the SHAP value of x_{ij}, indicating the contribution degree of the j-th feature in the i-th sample to the final predicted value: If $f(x_{ij}) > 0$, it indicates that this feature mainly plays a positive promoting role in the increase of the predicted value; conversely, if $f(x_{ij}) < 0$, it indicates that this feature mainly has a negative effect on the increase of the predicted value. The larger the absolute value of $f(x_{ij})$, the higher the degree of influence of this feature on the overall.

The prediction model of stainless-steel stress corrosion crack growth rate established in this research contains a total of three input features. Figure 8 shows the feature influence distribution map calculated by the SHAP method. From top to bottom, it indicates that the importance of the features gradually decreases. The color of the scatter points from blue to red represents that the feature value increases from small to large. The abscissa of the scatter points, the SHAP value, represents the positive or negative correlation between the feature value and the target value, and the absolute value indicates the contribution degree of the influence. It can be seen from the figure that the order of feature importance affecting the prediction results of the model from large to small is K, yield strength, and

temperature. All three features show a significant positive influence on the predicted value of crack growth rate. With the increase of the feature value, the SHAP value shows an upward trend.

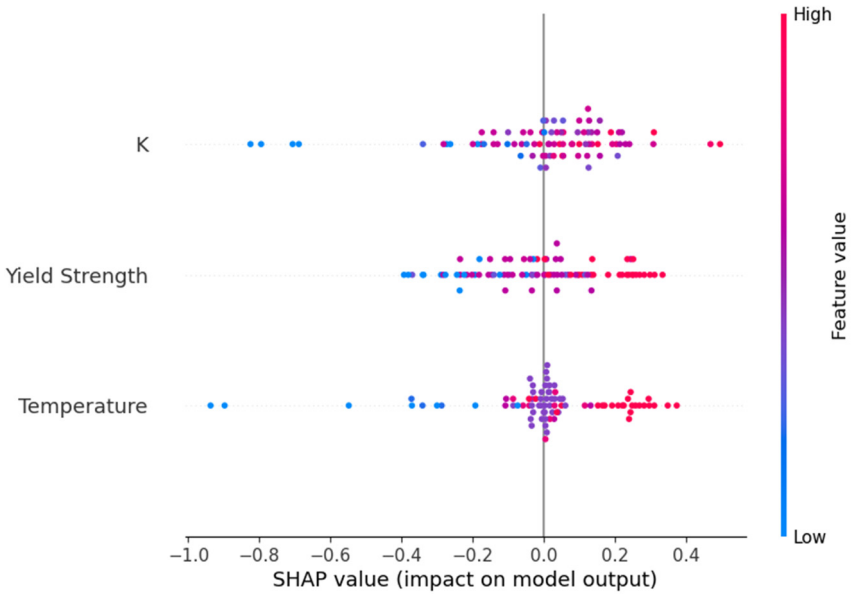


Figure 8. Feature influence distribution map.

To further explain the influence of each feature value on the crack growth rate, the prediction interpretation diagram of the SHAP value for a single sample is drawn, as shown in Figure 9. The left axis represents the feature name and feature value. The arrow direction of the bar graph indicates the influence direction of the feature value on the predicted value. The value represents the SHAP value of the feature value. E[f(x)] is the average value of the prediction of the entire data by the XGBoost model established in this paper, representing the baseline value of the entire model. The predicted result of the current sample can be obtained by accumulating the SHAP values of each feature value to the baseline value.

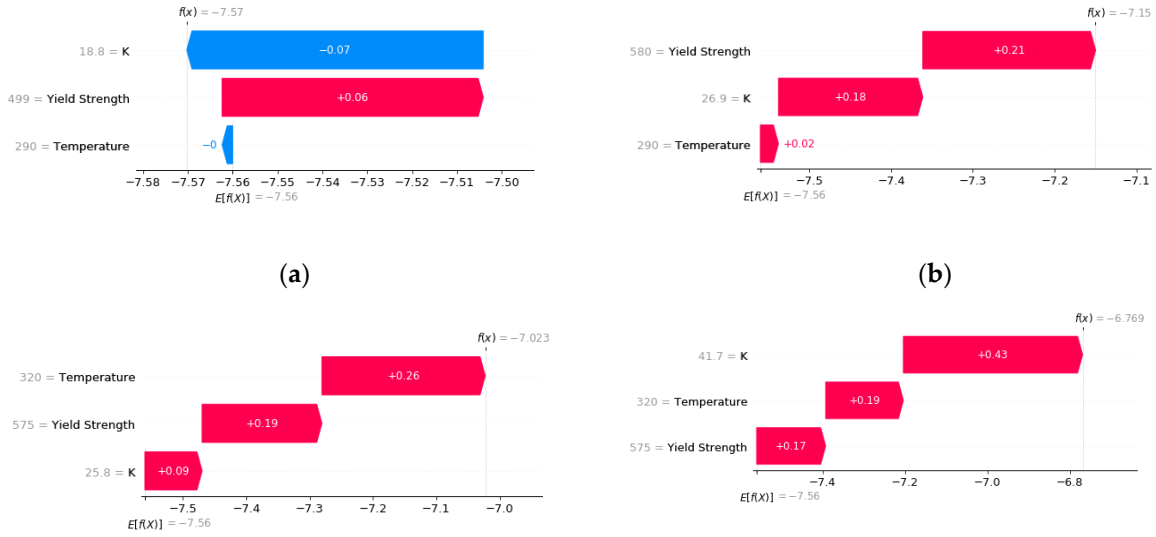


Figure 9. Prediction interpretation analysis of SHAP values for a single sample.

Taking the sample closest to the model's baseline value as the reference baseline (Figure 9a), when K increases to 23.6 and the yield strength rises to 580, the predicted value

of the crack growth rate increases to −7.388 compared to the baseline value (Figure 9b). K and the yield strength mainly contribute to the increase of the predicted value of the crack growth rate. Among them, the contribution value of the yield strength is 0.21 and that of K is 0.18. When the temperature further rises to 320 °C, the predicted value of the crack growth rate increases further (Figure 9c), and the temperature becomes the main factor promoting the growth of the crack growth rate, with a SHAP value of 0.26, followed by the yield strength and K, with SHAP values of 0.19 and 0.09 respectively. When K further increases to 41.7, the predicted value of the crack growth rate increases to −6.769 (Figure 9d), and K becomes the main factor promoting the growth of the crack growth rate, followed by the temperature and the yield strength, with SHAP values of 0.19 and 0.17 respectively.

The prediction interpretation and analysis of SHAP values for a single sample show that with the increase of K, yield strength, and temperature, the predicted value of crack growth rate by XGBoost increases. The influence degree of the feature values on the crack growth rate varies with the different increments of the features, indicating that the stainless-steel stress corrosion crack growth is a multi-factor competitive process, and the influence mechanism is consistent with the SCC mechanism. Through the global interpretation of feature importance and the local interpretation of feature contribution values, the prediction mechanism of XGBoost for the crack growth rate is revealed. It also indicates that the crack growth rate prediction model established in this study can accurately reflect the influence of each influencing factor on the SCC crack growth rate, further proving the reliability of the model's prediction results.

4. Conclusions

In this study, four machine learning algorithms, namely RF, XGBoost, SVR, and GPR, were utilized to establish the prediction model of the stainless-steel SCC crack growth rate in the simulated PWR primary water environment. The prediction accuracies of different models were compared and evaluated. The results show that XGBoost and RF, based on decision trees, have significant advantages in learning SCC crack growth rate data. Among them, XGBoost adds regularization terms and gradient boosting strategies to the cost function, which can simplify the model during the training process, enhance the generalization ability of the model, and effectively prevent overfitting. Therefore, XGBoost has the best prediction performance, with a corresponding R2 of 0.815, MSE of 0.061, MRE of 2.5%, and MAB of 0.124. Compared with the empirical equation, XGBoost possesses high flexibility and excellent data-driven learning ability. Ninety percent of the prediction errors in the test set are within the experimental values. The maximum error ratio is 2.5, and the average error ratio is 0.5. The prediction effect has been significantly enhanced. The interpretive analysis results of XGBoost based on the SHAP method indicate that the established model can accurately reflect the influence of each influencing factor on the SCC crack growth rate, further verifying the reliability of the prediction results. It holds significant reference value and application prospects for the prediction of stainless-steel stress corrosion behavior in the high-temperature and high-pressure water environment of PWR.

Author Contributions: Conceptualization, P.W.; methodology, P.W. and H.W.; validation, C.X. and X.L.; formal analysis, P.W.; investigation, P.W. and H.W.; resources, X.L.; data curation, C.X.; writing—original draft preparation, P.W.; writing—review and editing, X.L. and H.W.; visualization, C.X.; project administration, X.L.; funding acquisition, H.W. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Major projects on basic and applied basic research of Guangdong province (Grant No. 2019B030302011), the National Key Research and Development Program of China (Grant No. 2021YFB3702204, No. 2021YFA1600903), and the Guangdong Basic and Applied Basic Research Foundation (grant No. 2023B1515120082).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Training Database

Table A1. The database for stress corrosion crack growth rates in 316 and 304 stainless steels.

Cold Work (%)	Vickers Hardness	Carbon Content (mass%)	Temperature (°C)	K (Mpa$\sqrt{m}$)	Yield Strength (MPa)	Crack Growth Rate (mm/s)
15	254	0.047	270	36.9	509	1.20×10^{-7}
20	270	0.047	270	40.8	583	1.70×10^{-8}
5	184	0.047	300	34.0	259	4.20×10^{-9}
20	270	0.047	270	35.1	583	1.80×10^{-8}
15	243	0.04	290	19.0	446	1.80×10^{-8}
20	270	0.047	340	32.9	572	3.40×10^{-7}
5	184	0.047	290	34.1	260	6.30×10^{-9}
20	267	0.04	300	33.4	497	5.70×10^{-8}
20	267	0.04	280	31.2	500	6.50×10^{-9}
15	254	0.047	290	13.0	506	4.10×10^{-9}
10	214	0.04	320	31.1	383	2.90×10^{-8}
20	267	0.04	310	31.6	496	3.30×10^{-8}
15	243	0.04	310	33.0	443	2.40×10^{-8}
15	243	0.04	290	32.1	446	2.90×10^{-8}
5	184	0.047	320	26.1	256	5.50×10^{-9}
15	254	0.047	250	30.1	513	2.30×10^{-9}
10	219	0.047	330	31.7	376	1.30×10^{-8}
15	243	0.04	300	32.7	444	6.10×10^{-9}
20	267	0.04	340	32.8	491	1.50×10^{-7}
15	254	0.047	290	30.8	506	6.10×10^{-9}
20	267	0.04	290	32.9	499	3.80×10^{-8}
20	270	0.047	290	26.9	580	7.10×10^{-8}
10	219	0.047	320	27.9	377	1.80×10^{-8}
20	270	0.047	290	14.4	580	1.90×10^{-8}
15	254	0.047	290	13.5	506	2.00×10^{-8}
20	270	0.047	290	29.7	580	9.10×10^{-8}
20	267	0.04	290	18.5	499	2.40×10^{-8}
20	270	0.047	290	42.2	580	1.80×10^{-7}
15	254	0.047	320	39.7	501	5.00×10^{-8}
20	267	0.04	280	31.1	500	5.10×10^{-9}
20	270	0.047	290	38.6	580	3.80×10^{-8}
15	254	0.047	320	13.0	501	5.60×10^{-9}
15	243	0.04	320	31.5	441	7.70×10^{-8}
10	214	0.04	310	30.0	385	1.70×10^{-8}
10	219	0.047	320	34.6	377	4.40×10^{-8}
20	270	0.047	320	25.8	575	9.50×10^{-8}
20	270	0.047	290	24.0	580	5.10×10^{-8}
20	270	0.047	290	34.9	580	3.70×10^{-8}
20	270	0.047	320	31.1	575	1.10×10^{-7}
10	219	0.047	320	28.2	377	4.70×10^{-8}
15	254	0.047	290	25.7	506	1.40×10^{-8}
15	254	0.047	320	20.8	501	4.40×10^{-8}
10	219	0.047	320	25.8	377	3.60×10^{-8}
5	205	0.04	290	31.2	278	3.30×10^{-8}

Table A1. *Cont.*

Cold Work (%)	Vickers Hardness	Carbon Content (mass%)	Temperature (°C)	K (Mpa$\sqrt{m}$)	Yield Strength (MPa)	Crack Growth Rate (mm/s)
15	243	0.04	340	34.2	438	9.20×10^{-8}
20	267	0.04	310	31.4	496	2.90×10^{-8}
20	270	0.047	290	13.6	580	5.60×10^{-9}
20	270	0.047	320	21.6	575	6.50×10^{-8}
10	219	0.047	290	40.2	382	2.30×10^{-8}
20	270	0.047	320	29.6	575	1.10×10^{-7}
15	254	0.047	320	29.1	501	6.90×10^{-8}
20	267	0.04	310	31.8	496	1.00×10^{-7}
20	270	0.047	250	31.7	587	7.90×10^{-9}
20	267	0.04	290	18.8	499	2.70×10^{-8}
20	267	0.04	290	27.8	499	4.30×10^{-8}
15	254	0.047	290	21.9	506	2.00×10^{-8}
20	267	0.04	290	22.9	499	3.10×10^{-8}
20	270	0.047	310	34.4	576	4.00×10^{-8}
20	270	0.047	300	30.1	578	1.40×10^{-7}
10	214	0.04	290	25.1	388	1.40×10^{-8}
5	184	0.047	320	31.9	256	1.70×10^{-8}
10	219	0.047	310	30.2	379	9.60×10^{-9}
20	267	0.04	320	31.1	494	1.10×10^{-7}
10	214	0.04	290	30.5	388	3.80×10^{-8}
15	254	0.047	310	30.7	502	1.30×10^{-8}
15	243	0.04	290	23.1	446	6.00×10^{-8}
15	254	0.047	330	30.9	500	2.10×10^{-8}
10	214	0.04	280	31.6	389	1.00×10^{-8}
20	267	0.04	320	14.0	494	3.70×10^{-9}
20	270	0.047	320	39.3	575	1.40×10^{-7}
5	205	0.04	320	30.9	273	4.60×10^{-9}
15	243	0.04	330	31.8	440	2.40×10^{-8}
15	243	0.04	250	30.1	453	2.80×10^{-9}
5	184	0.047	320	33.9	256	1.50×10^{-8}
20	270	0.047	290	20.6	580	1.50×10^{-8}
20	270	0.047	360	32.1	570	1.40×10^{-7}
20	270	0.047	320	32.4	575	1.20×10^{-7}
20	270	0.047	290	40.2	580	3.40×10^{-8}
20	270	0.047	340	32.9	572	2.70×10^{-7}
20	267	0.04	290	31.0	499	1.90×10^{-8}
20	270	0.047	360	32.5	570	2.00×10^{-7}
20	267	0.04	270	33.3	502	1.60×10^{-8}
10	219	0.047	300	30.9	380	5.80×10^{-9}
15	254	0.047	290	25.7	506	1.60×10^{-8}
15	254	0.047	290	30.1	506	3.40×10^{-8}
15	254	0.047	320	25.3	501	4.50×10^{-8}
20	270	0.047	350	32.9	571	3.30×10^{-7}
20	267	0.04	290	23.6	499	4.10×10^{-8}
10	214	0.04	290	31.4	388	2.90×10^{-8}
15	254	0.047	290	17.3	506	2.20×10^{-8}
20	270	0.047	290	19.3	580	3.80×10^{-8}
20	270	0.047	320	41.7	575	1.70×10^{-7}
20	267	0.04	250	29.2	506	9.90×10^{-9}
15	254	0.047	290	30.4	506	1.70×10^{-8}
20	267	0.04	290	13.8	499	1.80×10^{-8}
20	270	0.047	290	26.6	580	5.00×10^{-8}
20	270	0.047	290	25.7	580	6.80×10^{-8}
20	267	0.04	320	30.5	494	8.70×10^{-8}

References

1. Huang, Y.; Wu, W.; Cong, S.; Ran, G.; Cen, D.; Li, N. Stress Corrosion Behaviors of 316LN Stainless Steel in a Simulated PWR Primary Water Environment. *Materials* **2018**, *11*, 1509. [CrossRef] [PubMed]
2. Zhang, L.; Chen, K.; Wang, J.; Guo, X.; Du, D.; Andresen, P.L. Effects of zinc injection on stress corrosion cracking of cold worked, austenitic stainless steel in high-temperature water environments. *Scr. Mater.* **2017**, *140*, 50–54. [CrossRef]
3. Wang, J.; Tianyu, Z.; Yaolei, H.; Jinna, M.B.; Fei, X.; Kai, C.; Du, D.; Andresen, P.L.; Zhang, L.; Zhang, M. Environment assisted cracking of 308L weld metal in high temperature water. *J. Nucl. Mater.* **2021**, *557*, 153275. [CrossRef]
4. Spencer, D.; Edwards, M.; Wenman, M.; Tsitsios, C.; Scatigno, G.; Chard-Tuckey, P. The initiation and propagation of chloride-induced transgranular stress-corrosion cracking (TGSCC) of 304L austenitic stainless steel under atmospheric conditions. *Corros. Sci.* **2014**, *88*, 76–88. [CrossRef]
5. Panda, B.; Sujata, M.; Madan, M.; Bhaumik, S. Stress corrosion cracking in 316L stainless steel bellows of a pressure safety valve. *Eng. Fail. Anal.* **2014**, *36*, 379–389. [CrossRef]
6. Couvant, T.; Legras, L.; Vaillant, F.; Boursier, J.M.; Rouillon, Y. Effect of strain-hardening on stress corrosion cracking of AISI 304L stainless steel in PWR primary environment at 360 °C. In Proceedings of the 12th International Conference on Environmental Degradation of Materials in Nuclear Power System, Salt Lake City, UT, USA, 14–18 August 2005; pp. 1069–1082.
7. Qu, H.J.; Wharry, J.P. Nanoindentation Investigation of Chloride-Induced Stress Corrosion Crack Propagation in an Austenitic Stainless Steel Weld. *Metals* **2022**, *12*, 1243. [CrossRef]
8. Li, Z.; Lu, Y.; Wang, X. Modeling of stress corrosion cracking growth rates for key structural materials of nuclear power plant. *J. Mater. Sci.* **2019**, *55*, 439–463. [CrossRef]
9. Andresen, P.L.; Ford, F.P. Life prediction by mechanistic modeling and system monitoring of environmental cracking of iron and nickel alloys in aqueous systems. *Mater. Sci. Eng. A Struct.* **1988**, *103*, 167–184. [CrossRef]
10. Shoji, T.; Lu, Z.; Murakami, H. Formulating stress corrosion cracking growth rates by combination of crack tip mechanics and crack tip oxidation kinetics. *Corros. Sci.* **2010**, *52*, 769–779. [CrossRef]
11. Macdonald, D.D.; Lu, P.; Urquidi-Macdonald, M.; Yeh, T. Theoretical Estimation of Crack Growth Rates in Type 304 Stainless Steel in Boiling-Water Reactor Coolant Environments. *Corrosion* **1996**, *52*, 768–785. [CrossRef]
12. Scott, P.M.; Calvar, M.L. *Sixth International Symposium on Environmental Degradation of Materials in Nuclear Power Systems—Water Reactors*; The Minerals, Metals & Materials Society: San Diego, CA, USA, 1993; p. 657.
13. Scott, P.M. *Proceedings of the Specialists Meeting on Operating Experience with Steam Generators, Brussels*; Committee on the Safety of Nuclear Installations: Brussels, Belgium, 1991; p. 5.
14. Hickling, J.; McIlree, A.; Pathania, R. *Materials Reliability Program (MRP): Crack Growth Rates for Evaluating Primary Water Stress Corrosion Cracking (PWSCC) of Thick-Wall Alloy 600 Materials (MRP-55) Revision 1006695*; EPRI: Palo Alto, CA, USA, 2002.
15. Seo, D.; Ogawa, K.; Nakao, Y.; Miura, H.; Shoji, T. Influence of high-temperature creep stress on growth of thermally grown oxide in thermal barrier coatings. *Surf. Coat. Technol.* **2009**, *203*, 1979–1983. [CrossRef]
16. Xue, H.; Wang, Z.W.; Zhao, L.Y.; Qiao, M.; Wang, W.B. Effect of Strength Matching of Welded Joint on Stress Corrosion Cracking in Nuclear Power Structures. *Appl. Mech. Mater.* **2013**, *457–458*, 593–596. [CrossRef]
17. Yamazaki, S.; Lu, Z.; Ito, Y.; Takeda, Y.; Shoji, T. The effect of prior deformation on stress corrosion cracking growth rates of Alloy 600 materials in a simulated pressurized water reactor primary water. *Corros. Sci.* **2008**, *50*, 835–846. [CrossRef]
18. Churyumov, A.Y.; Kazakova, A.A. Prediction of True Stress at Hot Deformation of High Manganese Steel by Artificial Neural Network Modeling. *Materials* **2023**, *16*, 1083. [CrossRef] [PubMed]
19. Li, R.; Deng, Q.; Tian, D.; Zhu, D.; Lin, B. Predicting Perovskite Performance with Multiple Machine-Learning Algorithms. *Crystals* **2021**, *11*, 818. [CrossRef]
20. Juan, Y.; Dai, Y.; Yang, Y.; Zhang, J. Accelerating materials discovery using machine learning. *J. Mater. Sci. Technol.* **2021**, *79*, 178–190. [CrossRef]
21. Mei, J.N.; Wang, P.; Han, Y.L.; Cai, Z.; Ti, W.X.; Peng, Q.J.; Xue, F. Prediction model of stress corrosion crack growth rate of nickel-based alloy 690 based on KBRF algorithm. *Rare Met. Mater. Eng.* **2022**, *51*, 1304–1311.
22. Rao, H.; Liu, D.; Jin, F.; Lv, N.; Nan, J.; Wang, H.; Yang, Y.; Wang, J. A Machine Learning Approach for Segmentation and Characterization of Microtextured Regions in a Near-α Titanium Alloy. *Crystals* **2023**, *13*, 1422. [CrossRef]
23. Arioka, K.; Yamada, T.; Terachi, T.; Chiba, G. Cold Work and Temperature Dependence of Stress Corrosion Crack Growth of Austenitic Stainless Steels in Hydrogenated and Oxygenated High-Temperature Water. *Corrosion* **2007**, *63*, 1114–1123. [CrossRef]
24. Lu, Z.P.; Shoji, T.; Takeda, Y. Effects of water chemistry on stress corrosion cracking of 316NG weld metals in high temperature water. *Corros. Eng. Sci. Technol.* **2014**, *50*, 41–48. [CrossRef]
25. Terachi, T.; Yamada, T.; Miyamoto, T.; Arioka, K. SCC growth behaviors of austenitic stainless steels in simulated PWR primary water. *J. Nucl. Mater.* **2012**, *426*, 59–70. [CrossRef]
26. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
27. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
28. Mei, J.N.; Wang, P.; Han, Y.L.; Cai, Z.; Peng, Q.J.; Xue, F. Prediction of stress corrosion crack growth rate of Ni-based 600 alloy based on TPE-XGBoost algorithm. *Rare Met. Mater. Eng.* **2021**, *50*, 2399–2408.

29. Chen, Z.B. Research on Application of Regression Least Squares Support Vector Machine on Performance Prediction of Hydraulic Excavator. *J. Control Sci. Eng.* **2014**, *2014*, 686130. [CrossRef]
30. Horvath, D.; Marcou, G.; Varnek, A.; Kayastha, S.; León, A.D.L.V.D.; Bajorath, J. Prediction of Activity Cliffs Using Condensed Graphs of Reaction Representations, Descriptor Recombination, Support Vector Machine Classification, and Support Vector Regression. *J. Chem. Inf. Model.* **2016**, *56*, 1631–1640. [CrossRef]
31. Lee, S.H. Optimization of Cold Metal Transfer-Based Wire Arc Additive Manufacturing Processes Using Gaussian Process Regression. *Metals* **2020**, *10*, 461. [CrossRef]
32. Biesiada, J.; Duch, W. Feature selection for high-dimensional data: A pearson redundancy based filter. *Soft Comput.* **2007**, *45*, 242–249.
33. Moshier, W.; Brown, C. Effect of cold work and processing orientation on stress corrosion cracking behavior of alloy 600. *Corrosion* **2000**, *5*, 307–320. [CrossRef]
34. Hou, J.; Peng, Q.J.; Shoji, T. Effects of cold working path on strain concentration, grain boundary microstructure and stress corrosuon cracking in Allo. *Corros. Sci.* **2011**, *53*, 2956–2962. [CrossRef]
35. Zhai, Z.; Toloczko, B.; Olszta, M.J.; Bruemmer, S.M. Stress corrosion crack initiation of alloy 600 in PWR primary water. *Corros. Sci.* **2017**, *123*, 76–87. [CrossRef]
36. Hammadi, R.A.; Yi, Y.; Zaki, W.; Cho, P.; Jang, C. Mechanical factors in primary water stress corrosion cracking of cold-worked stainless steel. *Nucl. Eng. Des.* **2016**, *301*, 24–31. [CrossRef]
37. Du, D.; Song, M.; Chen, K.; Zhang, L.; Andresen, P.L. Effect of deformation level and orientation on SCC of 316L stainless steel in simulated light water environments. *J. Nucl. Mater.* **2020**, *531*, 152038. [CrossRef]
38. Enhou, H.; Jian-qiu, Y.; XinJiang, W.; Wei, K. Corrosion mechanisms of stainless steel and nickel base alloys in high temperature high pressure water. *Acta Metall. Sin.* **2010**, *46*, 1379–1390.
39. Du, D.; Wang, J.; Chen, K.; Zhang, L.; Andresen, P.L. Environmentally assisted cracking of forged 316LN stainless steel and its weld in high temperature water. *Corros. Sci.* **2019**, *14*, 69–80. [CrossRef]
40. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 4768–4777.
41. Yan, L.; Diao, Y.; Lang, Z.; Gao, K. Corrosion rate prediction and influencing factors evaluation of low-alloy steels in marine atmosphere using machine learning approach. *Sci. Technol. Adv. Mater.* **2020**, *21*, 359–370. [CrossRef]
42. Jeon, J.; Seo, N.; Son, S.B.; Lee, S.-J.; Jung, M. Application of Machine Learning Algorithms and SHAP for Prediction and Feature Analysis of Tempered Martensite Hardness in Low-Alloy Steels. *Metals* **2021**, *11*, 1159. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.