Article

A Crack Propagation Method for Pipelines with Interacting Corrosion and Crack Defects

Mingjiang Xie [1], Yifei Wang [1], Weinan Xiong [2], Jianli Zhao [1] and Xianjun Pei [1,*]

1   School of Mechanical Engineering, Southeast University, Nanjing 211189, China;
   mingjiang@seu.edu.cn (M.X.); yifeiwang@seu.edu.cn (Y.W.); 220210314@seu.edu.cn (J.Z.)
2   Department of Information System, Dalian Naval Academy, Dalian 116018, China; xiongweinan6616@163.com
*   Correspondence: xpei@seu.edu.cn

Abstract: Corrosion and crack defects often exist at the same time in pipelines. The interaction impact between these defects could potentially affect the growth of the fatigue crack. In this paper, a crack propagation method is proposed for pipelines with interacting corrosion and crack defects. The finite element models are built to obtain the Stress Intensity Factors (SIFs) for fatigue crack. SIF interaction impact ratio is introduced to describe the interaction effect of corrosion on fatigue crack. Two approaches based on extreme gradient boosting (XGBoost) are proposed in this paper to predict the SIF interaction impact ratio at the deepest point of the crack defect for pipelines with interacting corrosion and crack defects. Crack size, corrosion size and the axial distance between these two defects are the factors that have an impact on the growth of the fatigue crack, and so they are considered as the input of XGBoost models. Based on the synthetic samples from finite element modeling, it has been proved that the proposed approaches can effectively predict the SIF interaction impact ratio with relatively high accuracy. The crack propagation models are built based on the proposed XGBoost models, Paris' law and corrosion growth model. Sensitivity analyses regarding corrosion initial depth and axial distance between defects are performed. The proposed method can support pipeline integrity management by linking the crack propagation model with corrosion size, crack size and the axial distance. The problem of how the interaction between corrosion and crack defects impacts crack defect growth is investigated.

Keywords: pipeline; fatigue crack; corrosion; stress intensity factor; finite element; XGBoost

1. Introduction

Pipelines are widely used to transport oil and gas products over long distances. Ensuring pipeline safety is a prerequisite for the transportation of fuels such as oil and natural gas. Researchers are committed to constructing more accurate and effective health management models and improving the integrity management system of pipelines. Researchers [1–3] summarized the existing models in the field of pipeline integrity management and pointed out that, although the current models consider the accuracy of inline inspection tools, they are still too ideal and challenging to accurately reflect the proper working conditions of the pipeline. Metal-loss corrosion defects are significant threats to pipeline integrity. Some researchers use stochastic processes to describe uncertainties associated with the degradation of wall thickness incurred by corrosion defects. Wang et al. [4] proposed a stochastic corrosion growth model using the geometric Brownian bridge process. Ossai et al. [5] used a non-homogeneous linear growth pure birth Markov model to predict the degradation of internal corrosion defects in oil and gas pipelines. Bazan and Beck [6] employed a Poisson square wave process to describe the corrosion growth rate and compared the proposed non-linear stochastic model with the linear corrosion growth model. Qin et al. [7] proposed a corrosion growth model based on Inverse Gaussian process and Markov Chain Monte Carlo simulation method. Pan et al. [8] also used Inverse Gaussian process to characterize

the degradation process of defects. Peng et al. [9] proposed a Bayesian framework of Inverse Gaussian process models. Remaining useful life of pipelines with multiple defects was predicted in refs. [10–12]. Although these corrosion growth models take multiple corrosion defects into account, they hardly consider the interacting effects among these defects, let alone the interactions between different types of defects.

There are a number of papers investigating pipelines with interacting corrosion defects. Benjamin et al. [13,14] presented a detailed literature review of pipelines with interacting corrosion defects and a database of corroded pipe tests. Amandi et al. [15] proposed a finite element model combined with a curve fitting method to estimate the remaining strength of pipelines with interacting corrosion defects. Sun and Cheng [16] also implemented a 3D finite element model to investigate mechano-electrochemical interaction of multiple longitudinal corrosion defects. Soares et al. [17] presented a model to analyze the integrity of pipelines with interacting corrosion defects under internal pressure and thermal stresses. Chen et al. [18] used a nonlinear finite element model to study the failure pressure of X80 pipelines with interacting corrosion defects. Kuppusamy et al. [19] investigated the effect of interaction of corrosion defects on the buckling strength of pipelines. Assessing and managing crack defects is also a vital part of pipeline integrity management. The remaining useful life prediction for pipelines with a single crack defect was conducted in refs. [20–22]. As for pipelines with interacting crack defects, Zhang et al. [23] presented a numerical model and fatigue simulations to analyze the fatigue behaviors. The corrosive environment will affect the growth of the crack, which is called Stress Corrosion Cracking (SCC). Hu et al. [24] applied the Monte Carlo method to predict and evaluate SCC. Lu et al. [25] established an SCC crack growth model in a high pH environment and verified it through experiments. Sekhar [26] summarized the effects of various crack interactions. This study shows that it is necessary to include the analysis of the interaction coupling between crack defects. These studies are all about the interaction between different defects of the same type. However, the exploration of interaction impact between different types of defects is still lacking in the existing literature.

In pipelines, common pipeline defects, such as crack and corrosion, exist at the same time. Specifically, there is an interacting effect between the fatigue crack and corrosion defect in the same pipeline segment. Pipeline corrosion will change the strength of the pipeline in the surrounding area. If the corrosion and crack defects are adjacent, a certain interaction coupling will occur and impact the Stress Intensity Factor (SIF) of the crack surface, thereby affecting the propagation of fatigue crack. Therefore, the crack propagation model that considers the interaction between these two types of defects is conducive to formulating more accurate detection and maintenance strategies. Motivated by this need, this paper plans to study the interacting effects of corrosion and crack defects on pipeline crack propagation.

In this paper, a method was developed for predicting the propagation of fatigue crack for pipelines with interacting corrosion and crack defects. Crack length, crack depth, corrosion length, corrosion depth and the axial distance between the crack and corrosion defects are all considered when developing this method. The finite element models are built to obtain the SIF values with and without considering the interaction impact between these defects. The powerful regression model, XGBoost [27], is applied in this paper to predict the SIF interaction impact ratio. With synthetic data from finite element analysis modeling, two approaches are provided to fit and predict the SIF interaction impact ratio at the deepest point of the crack defect, considering the interaction between corrosion and crack defects. The first one uses the data samples to directly fit and predict the SIF interaction impact ratio with a XGBoost model. As for the second one, it is an indirect prediction approach. It fits the SIF with and without considering the interaction impact, respectively. Therefore, two XGBoost models are acquired in this approach. The prediction results from these two models are utilized to calculate SIF interaction impact ratio. SIF interaction impact ratio is defined as the ratio of the SIF considering the interaction impact divided by the SIF without considering the interaction impact. With the proposed XGBoost

models and traditional crack and corrosion growth models, a crack propagation model is proposed for pipelines with interacting crack and corrosion defects, and simulation results are obtained for sensitivity analysis.

The novelty of this paper is three-fold: (1) it studies the interaction impact between different types of defects in pipelines, viz. crack and corrosion defects, depending upon the crack size, corrosion size and the axial distance between them; (2) it introduces SIF interaction impact ratio to describe the degree of the interaction impact and employs an advanced machine learning algorithm XGBoost to fit and predict the SIF interaction impact ratio; and (3) it proposes a method to predict the propagation of fatigue crack considering the interaction impact.

The rest of the paper is organized as follows. Section 2 presents the finite element analysis model for a pipeline with interacting corrosion and crack defects. Section 3 presents the proposed crack propagation model based on XGBoost. In Section 4, experimental results are obtained to analyze the interaction impact. Conclusions are presented in Section 5.

2. The Pipeline Finite Element Analysis Model

2.1. The FEA Model

In this section, the finite element software ANSYS® is used to model the pipeline with interacting fatigue crack and external corrosion defects. In the modeling process, the pipeline models with and without corrosion defects are established, respectively, to analyze the interaction impact of corrosion defect on crack propagation. The material of the modelled pipeline is API 5L X70. The outside diameter of the pipeline is set as 914.4 mm, and the wall thickness is 15.875 mm. The internal pressure is assumed to be 1 MPa for modeling. The fatigue crack is modeled as a semi-elliptical shape with a length of 15.2 mm and a crack depth in the range of 2 mm–12 mm. The SIF values corresponding to the deepest point and edge point can be obtained through stress analysis. The internal pressure of the pipeline is 1 MPa. At the same time, there are cuboid corrosion defects on the outer surface of the pipeline, and the axial distance from the crack center to the corrosion center moves from 150 mm to 500 mm. The depth of the corrosion defect is from 2 mm to 14 mm, with an increment of 1 mm each time. The geometric modeling of a corroded pipeline is shown in Figure 1, and the finite element model built in this paper is shown in Figure 2.
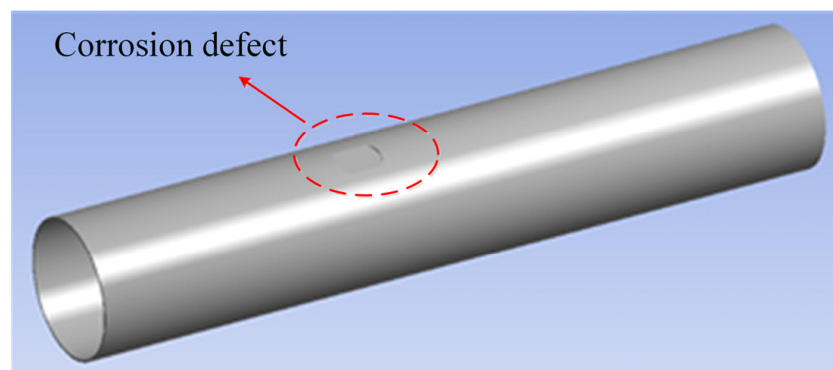

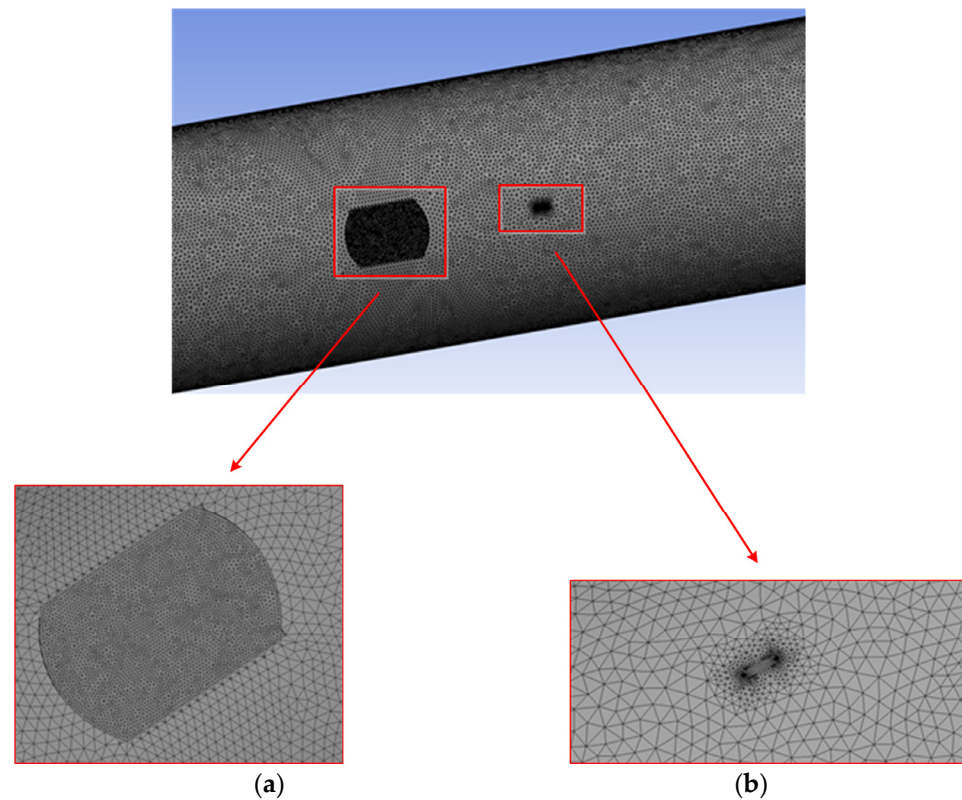**Figure 1.** Geometric modeling of corroded pipeline.

Figure 2. The grid division of pipeline. (**a**) Partial refinement of corrosion defect grid. (**b**) Partial refinement of grid at crack.

2.2. Validation of FEA

Generally, fracture in engineering structures can be classified into three types: opening mode (I), sliding mode (II) and tearing mode (III), and SIF is used to reflect these modes. Compared with mode II and III, SIF corresponding to mode I is much larger, so the mode I SIF dominates the propagation of fatigue crack. In this paper, mode I SIF was only considered in the pipeline remaining useful life prediction. The method based on API579 for the partial verification of FEA model was employed. According to API579 criterion, the SIF of mode I of the pipeline is calculated as follows:

$$K = \frac{pR_i^2}{R_o^2 - R_i^2} \left[2G_0 + 2G_1 \left(\frac{a}{R_o} \right) + 3G_2 \left(\frac{a}{R_o} \right)^2 + 4G_3 \left(\frac{a}{R_o} \right)^3 + 5G_4 \left(\frac{a}{R_o} \right)^4 \right] \sqrt{\frac{\pi a}{Q}} \quad (1)$$

$$Q = 1.0 + 1.464 \left(\frac{a}{c} \right)^{1.65}, \ for \ \frac{a}{c} \le 1 \quad (2)$$

$$G_0 = A_{0,0} + A_{1,0}\beta + A_{2,0}\beta^2 + A_{3,0}\beta^3 + A_{4,0}\beta^4 + A_{5,0}\beta^5 + A_{6,0}\beta^6 \quad (3)$$

$$G_1 = A_{0,1} + A_{1,1}\beta + A_{2,1}\beta^2 + A_{3,1}\beta^3 + A_{4,1}\beta^4 + A_{5,1}\beta^5 + A_{6,1}\beta^6 \quad (4)$$

$$\beta = \frac{2\phi}{\pi} \quad (5)$$

$$G_2 = \frac{\sqrt{2Q}}{\pi} \left(\frac{16}{15} + \frac{1}{3}M_1 + \frac{16}{105}M_2 + \frac{1}{12}M_3 \right) \quad (6)$$

$$G_3 = \frac{\sqrt{2Q}}{\pi} \left(\frac{32}{35} + \frac{1}{4}M_1 + \frac{32}{315}M_2 + \frac{1}{20}M_3 \right) \quad (7)$$

$$G_4 = \frac{\sqrt{2Q}}{\pi} \left(\frac{256}{315} + \frac{1}{5}M_1 + \frac{256}{3465}M_2 + \frac{1}{30}M_3 \right) \quad (8)$$

$$M_1 = \frac{2\pi}{\sqrt{2Q}}(3G_1 - G_0) - \frac{24}{5} \tag{9}$$

$$M_2 = 3 \tag{10}$$

$$M_3 = \frac{6\pi}{\sqrt{2Q}}(G_0 - G_1) + \frac{8}{5} \tag{11}$$

where p is the internal pressure; R_i is the internal radius; R_o is the outer radius; a is the crack depth; Q is a parameter based on crack geometry; G_0, G_1, G_2, G_3, G_4, M_1, M_2, M_3, $A_{i,j}$ ($i \in \{0,1,2,3,4,5,6\}$, $\{j \in 0,1\}$), β are influence coefficients; ϕ is the included angle; c is the half crack length; and K is the mode I SIF.

The finite element simulation results are compared with the SIF results calculated according to API 579 criterion. The results are shown in Figure 3. It can be found from the figure that for the pipeline without corrosion defects, the SIF obtained by finite element simulation is very close to the results of theoretical calculation for a large portion of crack depth range, and the maximum error is less than 5%. The accuracy of finite element simulation is proved. Then, the SIF of pipeline with corrosion defects is studied. As is obtained from Figure 3, for the same crack depth, the SIF of the pipeline with corrosion defects is greater than that without corrosion defects. With the increase in crack depth, SIF also increases gradually. The comparison results demonstrate that there is an interacting impact of corrosion and crack defects on SIF values. Therefore, it is necessary to study the interacting impact between corrosion and crack defects.
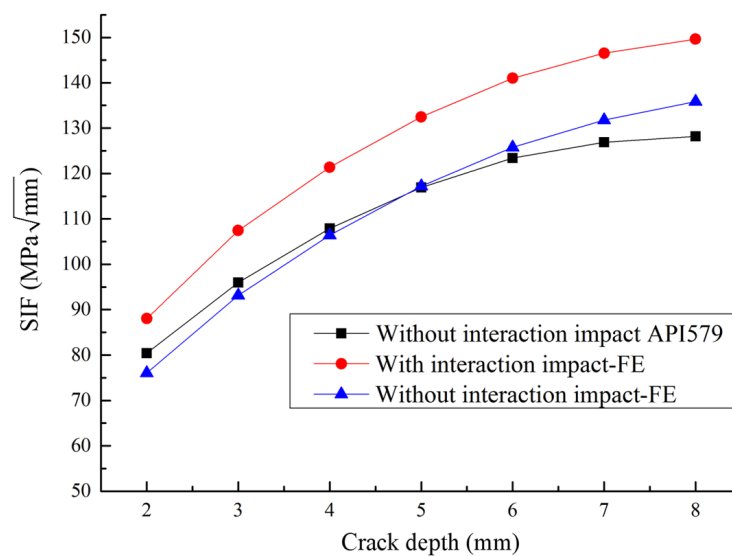

Figure 3. Comparison of pipeline SIF results.

3. The Proposed Crack Propagation Method Based on Extreme Gradient-Boosting Algorithm

3.1. The Extreme Gradient-Boosting Model

Extreme Gradient Boosting (XGBoost) is an ensemble machine learning algorithm based on Decision Tree and uses Gradient Boosting as the framework. It is developed from Gradient-Boosting Decision Tree (GBDT). GBDT is an additive model based on boosting, which is a general ensemble method. It employs a forward stagewise algorithm for greedy learning in the training process. In each iteration, GBDT learns a Classification and Regression Tree (CART), where Figure 4 is an example of CARTs, to fit the residual error between the prediction result from previous CARTs and the actual value of the training dataset. In other words, it is there to build a model from the training dataset and create a second model to correct the residual error from the first model. Then, the models are added until the training dataset is predicted relatively accurately, or a maximum number of models is added.
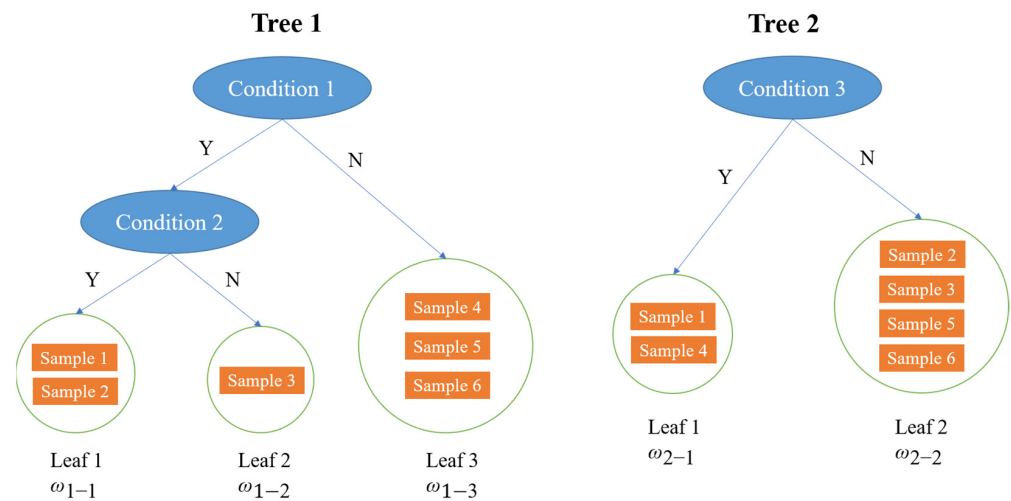
Tree 1   **Tree 2**

Figure 4. An example of CARTs.

Several optimization strategies are added into XGBoost model. Firstly, in order to improve computational accuracy, XGBoost uses the second-order derivative to optimize the objective function. Conversely, GBDT only uses the first-order derivative for optimization. In addition, the objective function of XGBoost utilizes regularization term to simplify the model and avoid overfitting. On the contrary, the GBDT does not have any regularization term in the objective function. XGBoost is able to automatically process default values and compute in parallel through a block storage structure, which cannot be implemented in GBDT. Since XGBoost has a high precision on the second-order derivative and fast parallel computation speed, it is very efficient in data processing and data modeling. In addition, XGBoost is relatively flexible, as it supports classification and regression, and it is able to provide customized objective function. XGBoost can be used with multiple programing languages and platforms. Therefore, XGBoost is widely used in the areas of data mining, recommender system and so on.

The objective function of XGBoost in the training process consists of two parts: loss function and regularization term:

$$Obj(\Theta) = L(\Theta) + \Omega(\Theta) \tag{12}$$

where Θ is the parameters obtained from the training processing; $L(\Theta)$ is the training error, which denotes the matching degree of the model to the training dataset; $\Omega(\Theta)$ is the regularization term, which represents the complexity of the model. Assuming that the training dataset is $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, the training error L can be expressed as the following equation:

$$L = \sum_{i=1}^{n} l(y_i, \hat{y}_i) \tag{13}$$

where y_i and $\hat{y}_i$ are the target output and the predicted output of the i-th sample x_i ($x_i \in R^z$, z is the number of features of the dataset), respectively, and n is the number of samples in the training dataset. For the proposed gradient-boosted machine, $l(y_i, \hat{y}_i) = (y_i - \hat{y}_i)^2$. The objective is to minimize $Obj(\Theta)$, which means $L(\Theta)$ and $\Omega(\Theta)$ should be relatively small. During the training process, it is required to balance the tradeoffs between bias and variance. Bias is controlled by $L(\Theta)$ and variance is controlled by $\Omega(\Theta)$. $L(\Theta)$ and $\Omega(\Theta)$ would be relatively large if underfitting. If overfitting, $\Omega(\Theta)$ would also be relatively large, since the model is weak on scalability and stability. Assuming there are V CARTs in the model, then

$$\hat{y}_i = \sum_{v=1}^{V} f_v(x_i), \; f_v \in F \tag{14}$$

where F is the function space of all the CARTs in the model. $f_v(x_i)$ represents the weight of the i-th sample falling on the leaf in the v-th tree. For the example in Figure 4, $f_1(\text{sample2}) = w_{1\text{-}1}$, $f_2(\text{sample2}) = w_{2\text{-}2}$, $f(\text{sample2}) = w_{1\text{-}1} + w_{2\text{-}2}$. Then, the model parameters that will be optimized from the training process are $\Theta = \{f_1, f_2, \ldots, f_V\}$, where f_v denotes the weight distribution of the samples falling on the leaf in the v-th tree. The objective function is shown in Equation (15):

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{v=1}^{V} \Omega(f_v) \tag{15}$$

Next, the objective function will be optimized in three steps. The first step is to use the second-order Taylor series expansion to optimize the loss function. The predicted values can also be expressed as

$$\hat{y}_i^{(u)} = \hat{y}_i^{(u-1)} + f_u(x_i) \tag{16}$$

which is the same as the expression of the GBDT. $\hat{y}_i^{(u)}$ is the predicted value of x_i in tree u after the i-th iteration. Then, the objective function after the i-th iteration can be represented using Equation (17):

$$Obj^{(u)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(u-1)} + f_u(x_i)) + \sum_{v=1}^{u} \Omega(f_v) \tag{17}$$

Using the second-order Taylor series expansion, the loss function becomes

$$\sum_{i=1}^{n} l(y_i, \hat{y}_i^{(u-1)} + f_u(x_i)) \approx \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{(u-1)}) + g_i f_u(x_i) + \frac{1}{2} h_i f_u{}^2(x_i)] \tag{18}$$

where

$$g_i = \frac{d(l(y_i, \hat{y}_i^{(u-1)}))}{d(\hat{y}^{(u-1)})} \tag{19}$$

$$h_i = \frac{\partial^2(l(y_i, \hat{y}_i^{(u-1)}))}{\partial(\hat{y}^{(u-1)})^2} \tag{20}$$

and the objective function is expressed in Equation (21):

$$Obj^{(u)} \approx \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{(u-1)}) + g_i f_u(x_i) + \frac{1}{2} h_i f_u{}^2(x_i)] + \sum_{v=1}^{u} \Omega(f_v) \tag{21}$$

The second step is to optimize the regularization term by expanding the regularization term and removing the constant term. Since forward calculation is adopted in XGBoost, then the structure of the $(u-1)$-th tree has been confirmed:

$$l(y_i, \hat{y}_i^{(u-1)}) = constant \tag{22}$$

$$\sum_{v=1}^{u} \Omega(f_v) = \Omega(f_u) + \sum_{v=1}^{u-1} \Omega(f_v) \tag{23}$$
$$= \Omega(f_u) + constant$$

Then the objective function is expressed as follows:

$$
\begin{aligned}
Obj^{(u)} &\approx \sum_{i=1}^{n} [l(y_i, \hat{y}_i^{(u-1)}) + g_i f_u(x_i) + \tfrac{1}{2} h_i f_u{}^2(x_i)] + \Omega(f_u) + constant \\
&= \sum_{i=1}^{n} [g_i f_u(x_i) + \tfrac{1}{2} h_i f_u{}^2(x_i)] + \Omega(f_u) + [\sum_{i=1}^{n} l(y_i, \hat{y}_i^{(u-1)}) + constant] \\
&= \sum_{i=1}^{n} [g_i f_u(x_i) + \tfrac{1}{2} h_i f_u{}^2(x_i)] + \Omega(f_u) + constant
\end{aligned}
\tag{24}
$$

After removing the constant term, the simplified objective function is

$$Obj^{(u)} \approx \sum_{i=1}^{n} [g_i f_u(x_i) + \frac{1}{2} h_i f_u^2(x_i)] + \Omega(f_u) \tag{25}$$

The last step of the optimization process is to merge the coefficients of the first-degree term and the quadratic term. Regarding the definition of a tree, the weight vector of leaves is set as $w \in R^T$ and the mapping relationship between the leaves (viz. the structure of the tree) is defined as $q : R^Z \rightarrow \{1, 2, 3, \ldots, T\}$ where T is the number of leaves in the tree. Then, $q(x)$ denotes the location of the leaf, for sample x. For the example in Figure 4, $q(\text{sample2}) = 1$ in tree 1, $q(\text{sample2}) = 2$ in tree 2. $f_t(x)$ can be represented by

$$f_u(x) = w_{q(x)} \tag{26}$$

Here, the number of leaves T and smoothness of leaf weight (viz. L2 norm of leaf weights) are used to describe the complexity of the tree, so

$$\Omega(f_u) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 \tag{27}$$

For the example in Figure 4, $\Omega(f_1) = \gamma 3 + \frac{1}{2} \lambda (w_{1-1}^2 + w_{1-2}^2 + w_{1-3}^2)$, $\Omega(f_2) = \gamma 2 + \frac{1}{2} \lambda (w_{2-1}^2 + w_{2-2}^2)$. $I_j = \{i \mid q(x_i) = j\}$ is the instance set in leaf j, $j = 1, 2, \ldots, T$. Grouping all the training samples based on leaves and utilizing Equations (26) and (27), then, the objective function is

$$Obj^{(u)} \approx \sum_{j=1}^{T} [(\sum_{i \in I_j} g_i)w_j + \frac{1}{2}(\sum_{i \in I_j} h_i + \lambda)w_j^2] + \gamma T \tag{28}$$

At last, merging the first-degree term and the quadratic term, then

$$Obj^{(u)} \approx \sum_{j=1}^{T} [G_j w_j + \frac{1}{2}(H_j + \lambda)w_j^2] + \gamma T \tag{29}$$

where

$$G_j = \sum_{i \in I_j} g_i, H_j = \sum_{i \in I_j} h_i \tag{30}$$

For each leaf j, the objective function is expressed as follows:

$$f(w_j) = G_j w_j + \frac{1}{2}(H_j + \lambda)w_j^2 \tag{31}$$

As the objective function of each leaf in the overall objective function is independent, then the overall objective function will achieve the minimum value when each leaf's objective function is minimized. The optimal solution of the quadratic function of one variable is

$$w_j^* = -\frac{G_j}{H_j + \lambda} \tag{32}$$

At this point, each leaf weight is optimized, and the overall objective function achieves its optimal value, viz. the minimum value:

$$Obj^* = -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j}{H_j + \lambda} + \gamma T \tag{33}$$

The structure of the tree is also the best at this time. The optimal objective functions of Figure 4 are shown in Figure 5. The fewer objective functions there are, the better the tree structures are.
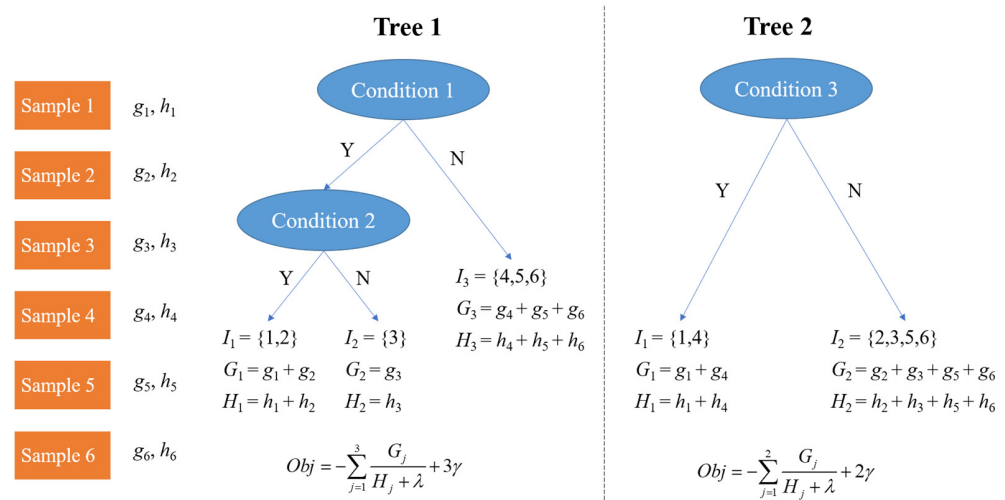

Figure 5. Objective function of the example in Figure 4.

In the actual training process, finding the optimal split point is a key problem. The applicable methods include greedy algorithm, approximate algorithm, weighted quantile sketch and sparsity-aware split finding. The greedy algorithm is the most commonly used.

3.2. The Proposed Model Based on XGBoost

In the proposed method, the scikit-learn wrapper interface for XGBoost was utilized to construct models to predict the SIF interaction impact ratio at the deepest point of the crack defect for pipelines with corrosion and crack defects. Based on the observations from finite element modeling, the size of crack and corrosion defects, and the axial distance between them, can affect SIF results. Therefore, the input variables of the proposed model are the length and depth of the crack defect, the length and depth of the corrosion defect, and the axial distance between the crack and the corrosion defects. The output variable is the interaction impact ratio α. The input and output variables are shown in Table 1. The crack length is assumed in the range of 15.2 mm–76.0 mm, and the crack depth is in the range of 2 mm–12 mm. The axial distance between the corrosion and crack defects is from 150 mm to 500 mm. The depth of the corrosion defect is from 2 mm to 14 mm.

Table 1. Input and output variables of the proposed XGBoost models.

Input Variables	Output Variables
Crack length	SIF considering interaction impact (K^*)
Crack depth	SIF without considering interaction impact (K)
Corrosion length	Interaction impact ratio (α)
Corrosion depth	
Axial distance between crack and corrosion defects	

In this paper, two approaches are provided to fit α. The first one is to directly construct a XGBoost model to predict α. The second one is to construct two XGBoost models to fit SIF values with and without considering the interaction impact, which are K^* and K, respectively. Then, the interaction impact ratio can be calculated with the formula $\alpha = K^*/K$. It is worth noting that only crack length and crack depth have an impact on SIF without considering interaction impact. Then, in the process of fitting K, there are only two input variables, viz. crack depth and crack length.

The samples used for modeling are synthetic data from finite element modeling. In total, 385 pieces of data are generated. A share of 80% of these data was randomly selected as the training set. The remaining data are the testing set. The scikit-learn API for XGBoost regression has a lot of parameters to set. In this paper, five parameters are selected for parameter tuning to get the best model structure and parameters: the number of gradient-boosted trees, the maximum depth of a tree, the minimum sum of instance weight needed in a child, $L1$ and $L2$ regularization terms on weights. The adjusting ranges for these five parameters are shown in Table 2. Increasing the maximum depth of a tree will make the model more complex, and it will be more likely to overfit, so the maximum value for this parameter is set to 10 in this paper. If the sum of instance weight in a leaf node is less than the minimum sum of instance weight needed in a child, the building process will stop further partitioning. Regarding the $L1$ and $L2$ regularization terms on weights, increasing their values will make the model more conservative. The learning rate is set at 0.1, which updates the weights to prevent overfitting and makes the boosting process more conservative. For the other parameters, such as the initial prediction score of all instances (global bias), minimum loss function required to make a further partition on a leaf node of the tree, etc., the default values in the scikit-learn API are applied.

Table 2. Parameter tuning for XGBoost models.

Parameters	Adjusting Ranges
Number of gradient-boosted trees	{40,50,60,70,80,90,100,110}
Maximum depth of a tree	{3,4,5,6,7,8,9,10}
Minimum sum of instance weight needed in a child	{1,2,3,4,5,6}
$L1$ regularization term on weights	{0.05,0.1,1,2,3}
$L2$ regularization term on weights	{0.05,0.1,1,2,3}

In the training process, a grid search method with 5-fold cross-validation was applied to select the best combination of the tuning parameters based on the determination coefficient R^2, which describes the goodness of fit of the current trained model. In other words, the original training set was re-segmented into the training set and validation set with the ratio of 4:1 five times, as shown in Figure 6. For each combination of the tuning parameters, the training set was used to train the model, and the validation set was used to evaluate the model's performance five times and compute the average performance, viz. average R^2, with these five times' results. This method can reduce training bias and improve the model's stability. After all the combinations' results are obtained, the model with the highest R^2 has the best combination of the tuning parameters. The values of R^2 are between 0 and 1. A value much closer to 1 indicates the regression model has a higher fitting degree.

In the actual training process, a pipeline of transforms with a final estimator (viz. model to be fitted) is utilized. This method is to sequentially apply a list of transforms and a final model. Intermediate steps of the pipeline must implement fit and transform methods, while the final model only needs to implement the fit method. In this paper, a pipeline consisting of a standard scaler and an XGBoost model is applied. The standard scaler is to normalize data to make its features have zero mean and unit variance. The standard scaler fits to the training set and transforms the training set and validation set.

The overall process of the first approach for constructing the XGBoost model is as follows:

Step 1. Randomly split the samples (output variable is SIF interaction impact ratio α) into training set and testing set with the ratio 8:2.

Step 2. Employ a pipeline consisting of a standard scaler and an XBoost model to the original training set for training. In detail, the 5-fold cross-validated grid search method is applied to the original training set to select the best model structure and parameters among all the combinations of the tuning parameters. The model with the highest R^2 is the best

model. The best model is then saved and can be directly applied to new data to acquire prediction values.

Step 3. Feed the testing set to the trained model to obtain the value of R^2, which indicates the ability fitting to new data with the trained model. The closer that R^2 is to 1, the better structured the model is. If the value is close to 1, then the trained model can be used to directly predict interaction impact ratio α.
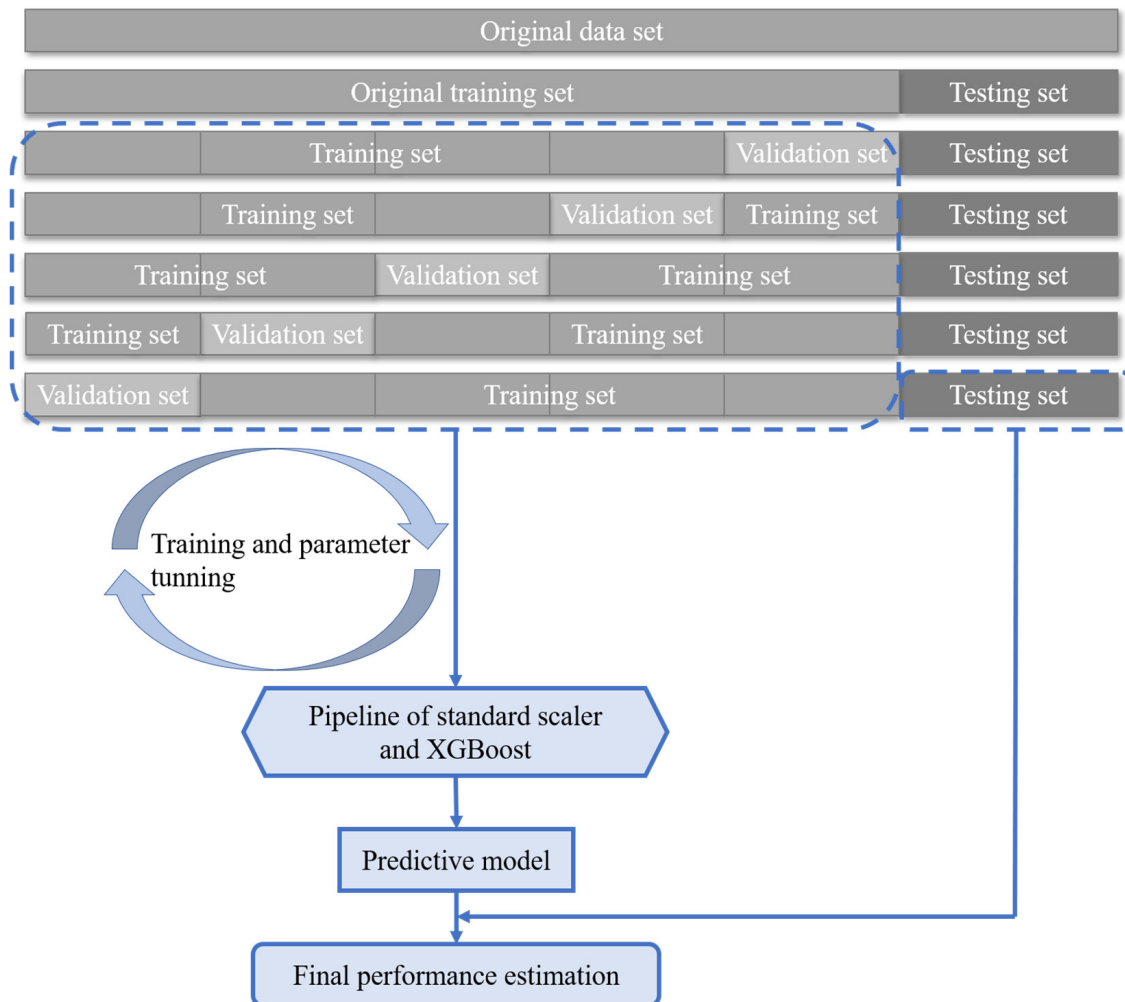

Figure 6. The procedure of the proposed algorithm.

Similarly, the overall process of the second approach for constructing the two XGBoost models is:

Step 1. Randomly split the samples (output variables are SIF values with and without considering interaction impact, viz. K^* and K) into training set and testing set with the ratio 8:2.

Step 2. When the output variable is K, the input's variables are crack depth and crack length, and employ a pipeline of a standard scaler and an XGBoost model to the original training set for training. In the same way, the 5-fold cross-validated grid search method is applied to the original training set to select the best model. The best model is saved to predict K.

Step 3. When the output variable is K^*, the input includes all the five input variables and employs a pipeline of a standard scaler and an XGBoost model to the original training set for training. In the same way, the 5-fold cross-validated grid search method is applied to the original training set to select the best model. The best model is then saved to predict K^*.

Step 4. Respectively, feed the two testing sets to the two trained models to obtain the values of R^2. If the two values of R^2 are close to 1, then the two trained models can be used to predict SIF with and without considering interaction impact, respectively.

Step 5. Calculate predicted interaction impact ratio α on the testing set with predicted K and K^* and compare the predicted values with the target ratio values by calculating R^2.

3.3. The Pipeline Corrosion and Fatigue Crack Growth Models

In the proposed model, corrosion defect is assumed to grow linearly. The growth of the corrosion depth is characterized by

$$d(t) = d_0 + g_d t \tag{34}$$

where d_0 represents the corrosion initial depth, g_d is the growth rate of corrosion depth, and t is the propagation time. The corrosion depth is used as the input variable in the XGBoost model to calculate the SIF interaction impact ratio. In this paper, the corrosion depth growth rate is assumed to be 0.3 mm/year [11].

Pipeline fatigue crack growth is predicted using the physics-based methods governed by Paris' law, which was employed in [28–30]. Based on Paris' law and the proposed model for evaluating the SIF interaction impact between corrosion and crack defects, the fatigue crack growth model is introduced in the following equation:

$$da/dN = C(\Delta K \alpha)^m \tag{35}$$

where da/dN is crack growth rate; a is crack depth; N is the number of loading cycles; α is the SIF interaction impact ratio; and ΔK is the range of SIF. C and m are material-related model parameters, which can be estimated via experiments. In this paper, it is assumed that model parameters $C = 5 \times 10^{-12}$, $m = 3$ [21]. Methods based on FE and XGBoost models are employed to calculate the SIF and SIF interaction impact ratio at the deepest point of the fatigue crack. In this study, this paper focuses on the crack depth growth, since the length is mostly unchanged.

4. Results

When directly fitting the SIF interaction impact ratio, the average determination coefficient on the validation sets during cross validation is 0.9935, and the standard deviation is 0.0041. Thus, it can be seen that the trained model has a relatively high stability. The prediction result on the testing set is as Figure 7 shows. On the testing set, the determination coefficient R^2 is 0.9876, which means the developed model can accurately predict the SIF interaction impact ratio. At this point, the number of gradient-boosted trees is 110, the maximum depth of a tree is 6, the minimum sum of instance weight needed in a child is 1, and the $L1$ and $L2$ regularization terms on weights are 0.05 and 0.1, respectively.

The prediction results of the SIF interaction impact ratio are shown in Figure 8. As observed in Figure 8, it can be found that the interaction impact ratio decreases as the crack depth a increases. From Figure 8a to Figure 8c, as corrosion depth increases from 4 mm to 10 mm, the SIF interaction impact ratio increases a lot when the axial distance between two corrosion and crack defects remains the same. Thus, the corrosion depth does affect the SIF interaction impact ratio a lot. The highest SIF interaction impact ratio in Figure 8c is 1.1848, which means it is necessary to consider the interaction impact between these two defects in the crack propagation process. The comparison results for different corrosion depths when crack depth is equal to 6 mm are shown in Figure 8d. From all these four figures, it can be found that the SIF interaction impact ratio is overall decreasing as the axial distance increases. These ratios first decrease quickly when the axial distance is smaller than around 175 mm and then decrease relatively slowly when the axial distance is in the range of 175 mm and 240 mm. When the axial distance is bigger than 240 mm, the decreasing speed is getting even smaller. This is because the corrosion defect moves away from the stress concentration zone of the crack defect.
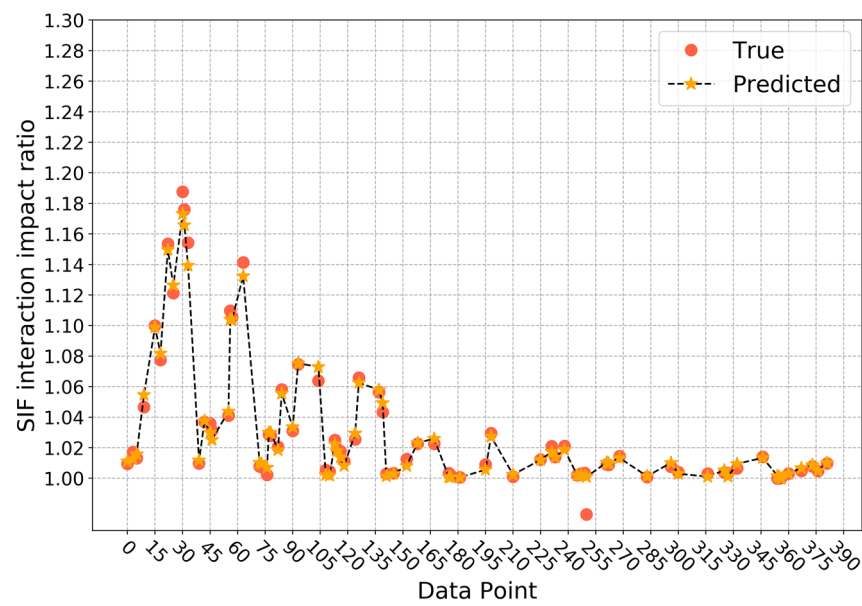
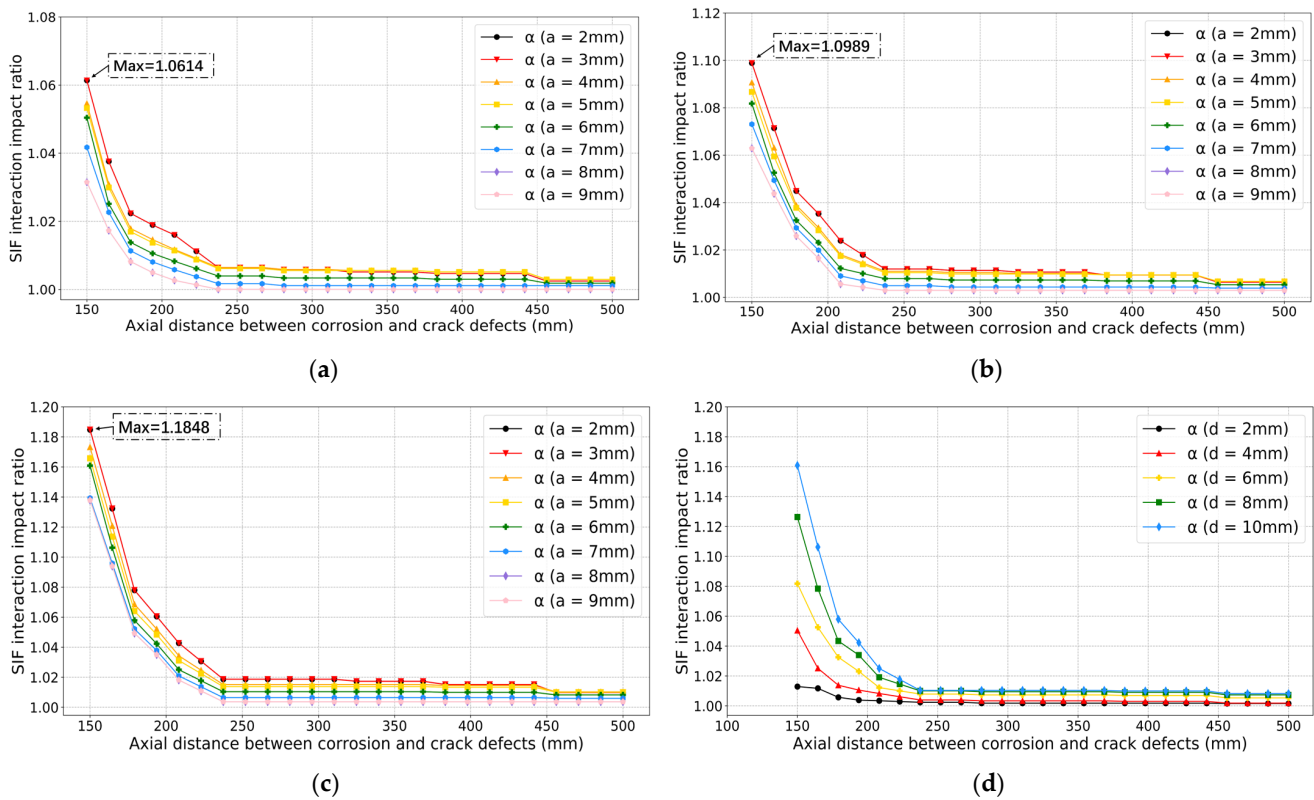**Figure 7.** Prediction results of SIF interaction impact ratio based on approach 1.

Figure 8. Prediction results of SIF interaction impact ratio based on approach 1. Corrosion depth *d*: (**a**) 4 mm; (**b**) 6 mm; (**c**) 10 mm; (**d**) Comparison results for different corrosion depths when crack depth *a* is 6 mm.

For the second approach, the average determination coefficient on the validation sets is 1.0000 and the standard deviation is 0.0000 when predicting SIF values without considering interaction impact, which means the trained model is relatively stable. Here, the number of gradient-boosted trees is 110, the maximum depth of a tree is 3, the minimum sum of instance weight needed in a child is 1, and the *L*1 and *L*2 regularization terms on weights are both 0.05. The prediction result on the testing set is shown in Figure 9, where the

determination coefficient R^2 is 1.0000. When considering interaction impact, the average determination coefficient on the validation sets is still 0.9998, and the standard deviation is still 0.0001. However, the selected structure and parameters of the model are different. The number of gradient-boosted trees is 110, the maximum depth of a tree is 4, the minimum sum of instance needed in a child is 2, and the $L1$ and $L2$ regularization terms on weights are 0.1 and 1, respectively. On the testing set, the determination coefficient R^2 is 0.9992, and the prediction result is as displayed in Figure 10. It can be seen that for these two predictive models, the performance is quite stable on the validation sets and very accurate on the testing set. Therefore, it can be concluded that these two XGBoost models can predict SIF values with and without considering interaction impact efficiently and accurately. Furthermore, this indicates these two models are able to predict interaction impact ratio efficiently and accurately, since the ratio is calculated from the predicted results of these two models. After the predictive results are obtained from these two models, the SIF interaction impact ratio α can be calculated with the equation K^*/K. For the testing set in this paper, the result is shown in Figure 11. At this time, the determination coefficient R^2 of the SIF interaction impact ratio on the testing set is 0.9852. From the experimental result, it can also be concluded that the two trained XGBoost models can predict SIF interaction impact ratio at the deepest point of the crack defect, considering the interaction between corrosion and crack defects accurately and efficiently.
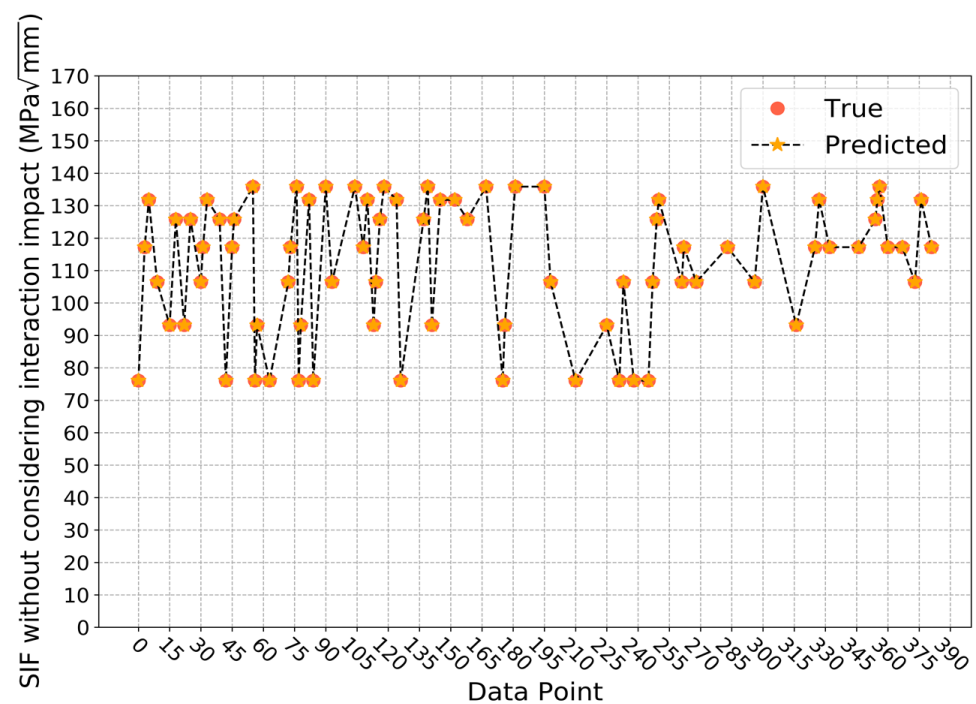

Figure 9. Prediction results of SIF values without considering interaction impact based on approach 2.

The comparison results of SIF values with and without considering interaction impact are shown in Figure 12. When increasing the crack depth a from 2 mm to 9 mm, K^* and K values both gradually increase as expected. From Figure 12a to Figure 12c, K^* values gradually increase, while K values remain the same as the corrosion depths increases. From the observations of these in Figure 12d, K^* and K values have relatively big differences when the axial distance between two defects is smaller than around 240 mm.
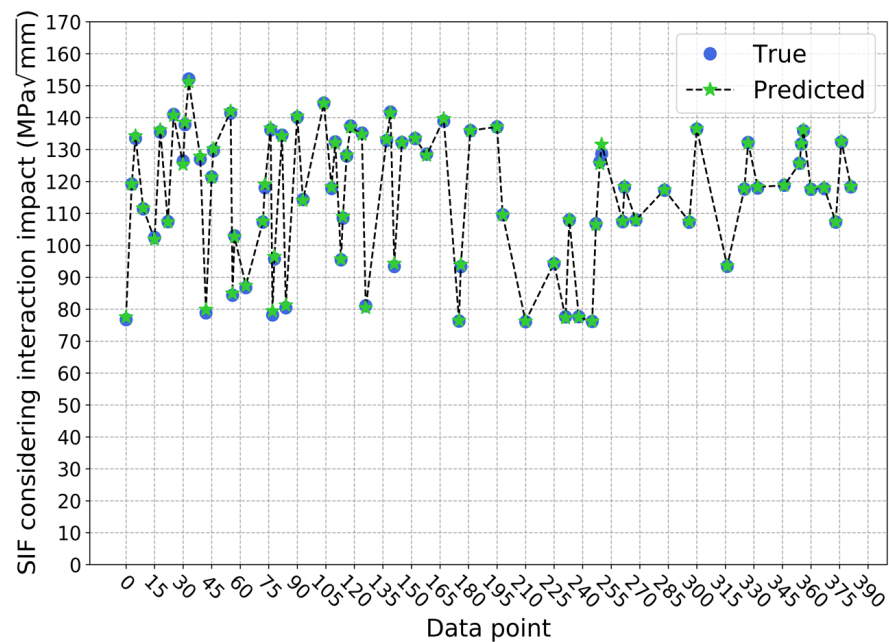
Figure 10. Prediction results of SIF values considering interaction impact based on approach 2.
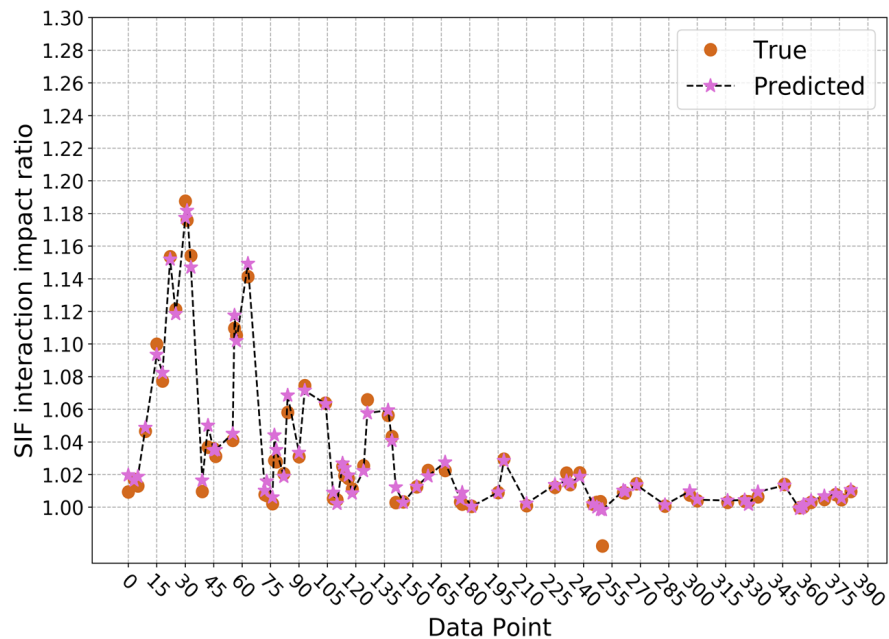

Figure 11. Prediction results of SIF interaction impact ratio based on approach 2.

The crack propagation models are built based on the proposed XGBoost model and Paris' law. The crack initial depth is set at 2 mm, since SIF interaction impact ratio is relatively large when crack depth is small. The corrosion initial depth is assumed at 6 mm. Figure 13a–d show the comparison results of crack depth growth models for different axial distances between two defects using approach 1 and 2, respectively. The red dash lines represent the crack critical depth, which is approximately 80% of the wall thickness. When the crack depth exceeds the crack critical depth, it is considered a failure. The comparison results shown in Figure 13 indicate that the crack depth predicted by approach 2 reaches the threshold more quickly than approach 1. The comparison results of the crack depth propagation curves for different axial distances based on approach 2 are shown in Figure 14. If the interaction impact between two defects is not considered, it takes about 14.8 years to fail. Meanwhile, when considering the interaction impact between two defects by

implementing the proposed model, the failure time changes from 9.2 to 13.8, 14.0 and 14.5 years, as the axial distance changes from 150 to 200, 300 and 350 mm. There is a big difference between the two crack propagation curves in Figure 13a, since the corrosion defect is in the stress concentration zone (axial distance smaller than 175 mm).
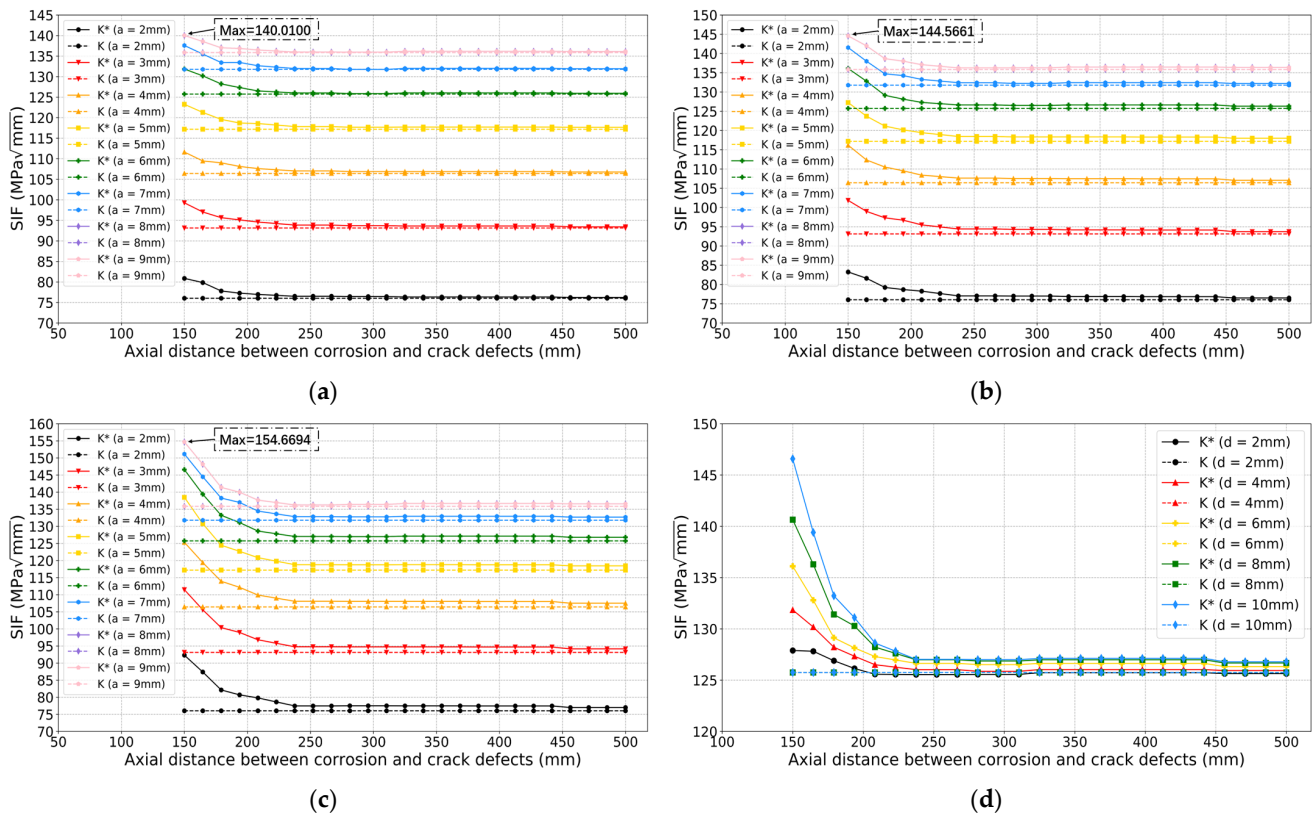

Figure 12. The comparison of SIF results based on approach 2. Corrosion depth *d*: (**a**) 4 mm; (**b**) 6 mm; (**c**) 10 mm; (**d**) Comparison results for different corrosion depths when crack depth *a* is 6 mm.

To perform sensitivity analysis regarding corrosion initial depth, the aggressive case was studied, in which the axial distance was set at 150 mm. The corrosion initial depth varies from 2 mm to 8 mm. Figure 15a–d show the results of crack depth growth models for different corrosion initial depths. These figures indicate that the crack depth grows more quickly using approach 2 than approach 1. The comparison results of the crack depth propagation curves for different corrosion initial depths based on approach 2 are shown in Figure 16. If not considering the interaction impact between two defects, it also takes about 14.8 years to fail. From the comparison results in Figures 15 and 16, the time to reach critical crack depth is 10.8, 9.9, 9.2 and 8.8 years, respectively, as the corrosion initial depth changes from 2 to 3, 6 and 8 mm. From the experimental results obtained from Figures 13–16, it can be concluded that the interaction impact between corrosion and crack defects affects the propagation of fatigue crack a lot. Thus, it is necessary to consider the SIF interaction impact ratio in the remaining useful life prediction, especially when the corrosion defect is in the stress concentration zone.
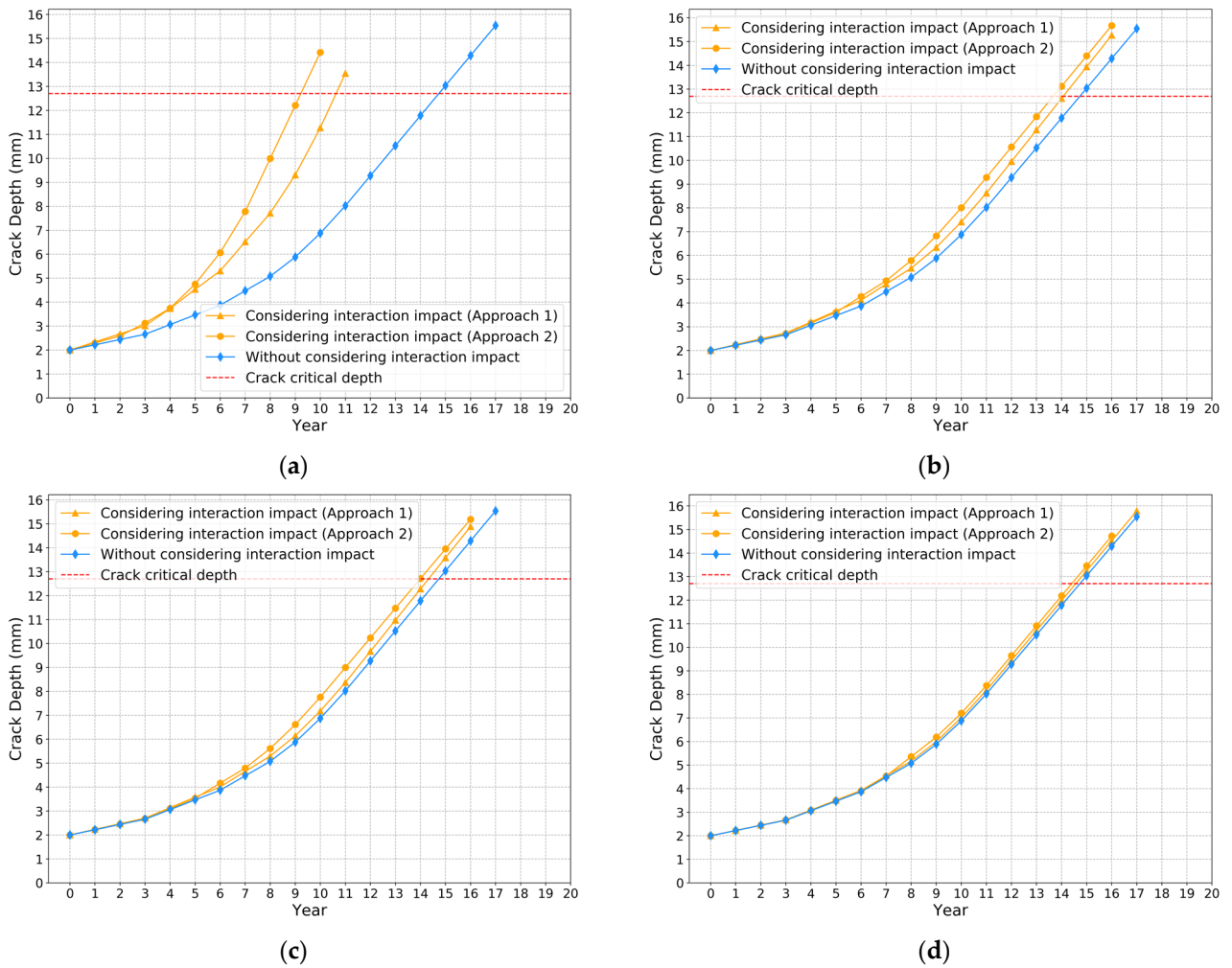
Figure 13. Investigations of the interaction impact on crack depth growth for different axial distances. Axial distance: (**a**) 150 mm; (**b**) 200 mm; (**c**) 300 mm; (**d**) 500 mm.
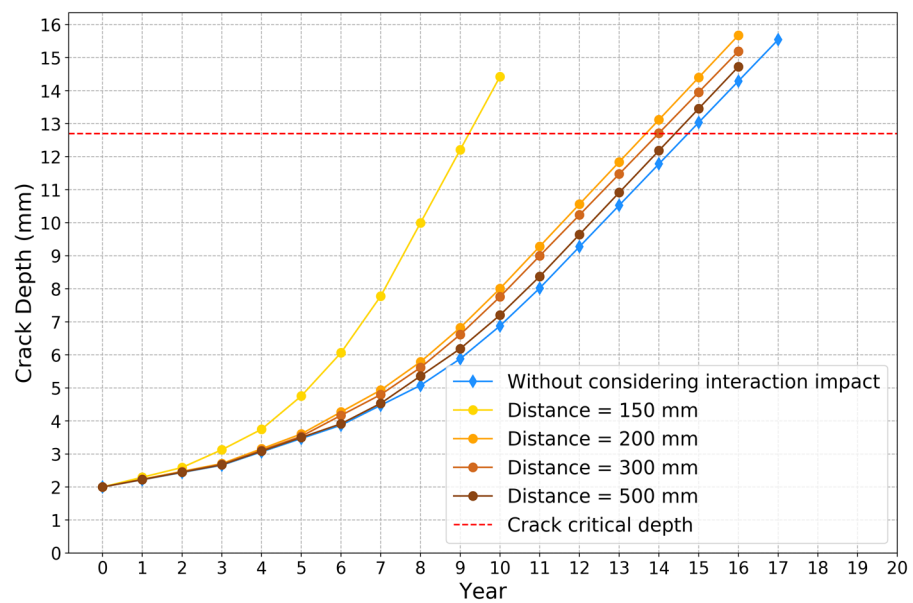

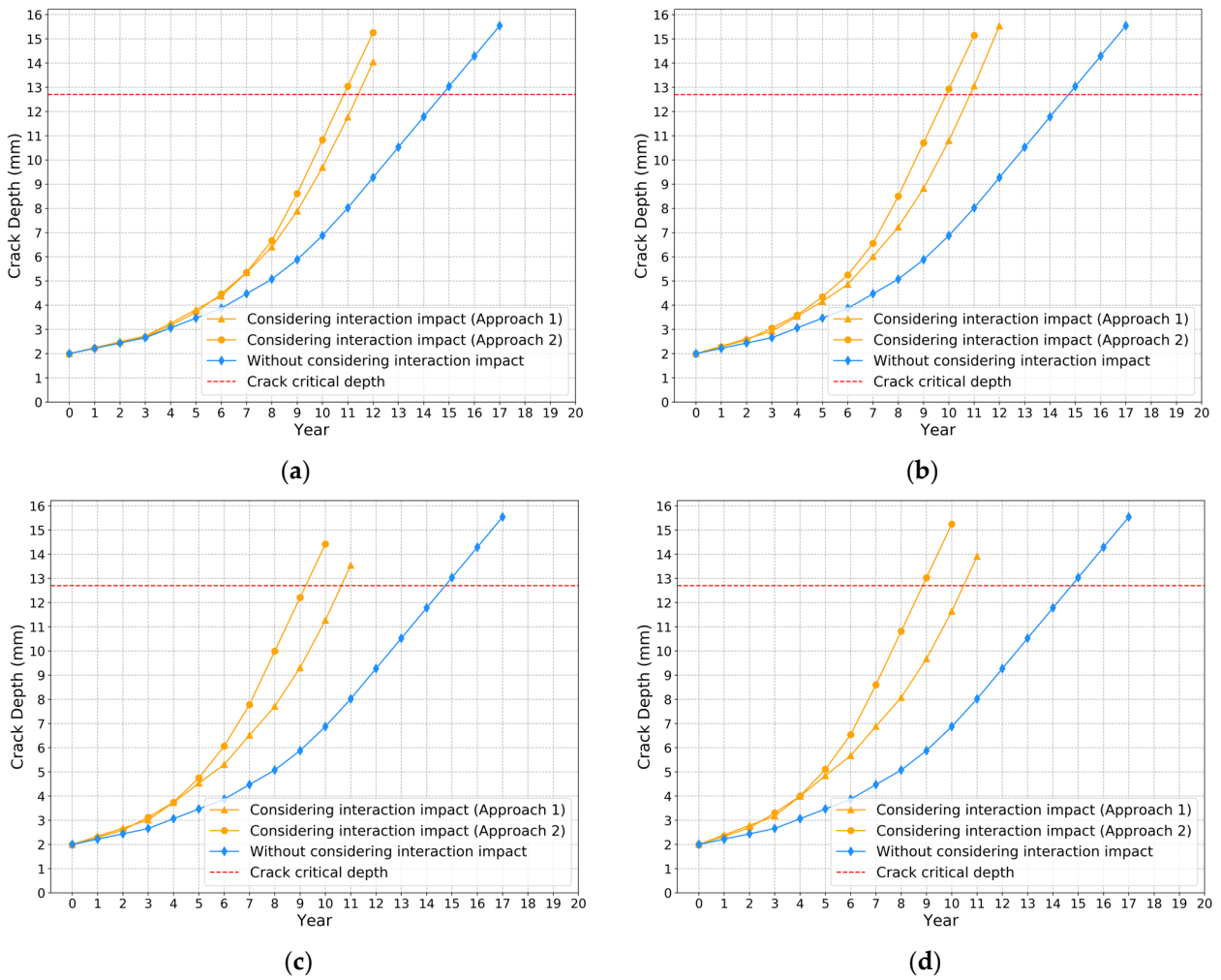**Figure 14.** Crack depth growth curves for different distances based on approach 2.

Figure 15. Investigations of the interaction impact on crack depth growth. Corrosion initial depth: (**a**) 2 mm; (**b**) 4 mm; (**c**) 6 mm; (**d**) 8 mm.
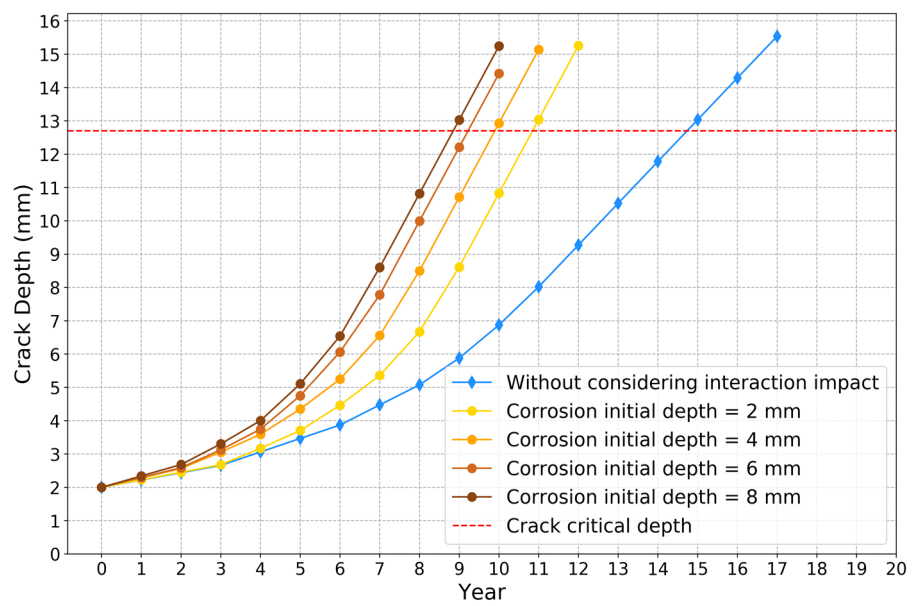

Figure 16. Crack depth growth curves for different corrosion initial depths based on approach 2.

5. Conclusions

The existing reported work only focuses on pipeline life prediction with single or multiple defects of the same type. The interaction impacts between different types of defects are not considered. In this work, the interaction impacts between crack and corrosion defects were studied, and a fatigue crack propagation method considering these impacts was proposed based on XGBoost models and Paris' law. Crack size, corrosion size, and the axial distance between these two defects were all considered in the proposed method. In addition, this paper introduced SIF interaction impact ratio to describe how the corrosion defect affects the stress concentration zone of the fatigue crack. Two approaches were implemented for SIF interaction impact ratio prediction. The first one directly fitted and predicted SIF interaction impact ratio with the synthetic samples from finite element modeling. The second one fitted and predicted the SIF with and without considering interaction impacts, respectively, and then calculated the SIF interaction impact ratio. Examples were used to demonstrate the proposed method. The determination coefficients of these two approaches on the testing sets were 0.9876 and 0.9852, respectively, which was quite close to 1. Therefore, it can be concluded that the developed method can predict fatigue crack growth accurately. Several key findings are listed below:

The SIF interaction impact ratio decreases as the crack depth increases. It increases as the corrosion depth increases.

The SIF interaction impact ratio is gradually decreasing as the axial distance increases. This ratio is relatively large when the axial distance is smaller than 240 mm.

The time to reach critical crack depth decreases as the corrosion initial depth increases or the axial distance decreases.

The method developed in this paper can support the decision making in pipeline integrity planning, especially when the corrosion defect is relatively close to the crack defect. However, the proposed method only considered the interacting impact between two defects. More efficient crack and corrosion propagation models considering more than two defects are desired in future research. Another research topic is to develop crack propagation models for different types of crack shapes instead of semi-elliptical shapes.

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

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Nomenclature

p	pipeline internal pressure
a	crack depth
c	half crack length
R_i	pipeline internal radius
R_o	pipeline outer radius
Q	crack geometry parameter in API 579

$G_0, G_1, G_2, G_3, G_4,$ $M_1, M_2, M_3, A_{i,j}$ $(i \in \{0,1,2,3,4,5,6\},$ $j \in \{0,1\}), \beta$	influence coefficients in API 579
ϕ	included angle in API 579
Obj	objective function of XGBoost
Θ	parameters obtained from the training processing in XGBoost
L	training error in XGBoost
Ω	regularization term in XGBoost
S	training dataset
l	training error of each sample in XGBoost
x_i	i-th sample
y_i	target output of the i-th sample
$\hat{y}_i$	predicted output of the i-th sample
z	number of features in the dataset
n	number of samples in the training dataset
V	number of classification and regression trees in XGBoost
v	v-th classification and regression tree in XGBoost
f_v	weights of samples falling on the leaf in the v-th tree
F	function space of all the classification and regression trees in XGBoost
g_i	first-order derivative of training error for i-th sample in XGBoost
h_i	second-order derivative of training error for i-th sample in XGBoost
w	weight vector of leaves in classification and regression tree
T	number of leaves in classification and regression tree
q	mapping relationship between the leaves in classification and regression tree (viz. the structure of the tree)
γ	coefficient for number of leaves in regularization term in XGBoost
λ	coefficient for L2 norm of leaf weights in regularization term in XGBoost
I_j	instance set in leaf j
G_j	sum of first-order derivatives of training error for leave j in XGBoost
H_j	sum of second-order derivatives of training error for leave j in XGBoost
K	SIF without considering interaction impact
K^*	SIF considering interaction impact
α	interaction impact ratio
d	corrosion depth
d_0	corrosion initial depth
g_d	growth rate of corrosion depth
t	propagation time
m, C	material parameters in Paris' law
N	loading cycles
ΔK	the range of SIF

References

1.   Vanaei, H.R.; Eslami, A.; Egbewande, A. A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models. *Int. J. Press. Vessel. Pip.* **2017**, *149*, 43–54. [CrossRef]
2.   Kishawy, H.A.; Gabbar, H.A. Review of pipeline integrity management practices. *Int. J. Press. Vessel. Pip.* **2010**, *87*, 373–380. [CrossRef]
3.   Xie, M.; Tian, Z. A review on pipeline integrity management utilizing in-line inspection data. *Eng. Fail. Anal.* **2018**, *92*, 222–239. [CrossRef]
4.   Wang, H.; Yajima, A.; Castaneda, H. A stochastic defect growth model for reliability assessment of corroded underground pipelines. *Process. Saf. Environ. Prot.* **2019**, *123*, 179–189. [CrossRef]
5.   Ossai, C.I.; Boswell, B.; Davies, I. Markov chain modelling for time evolution of internal pitting corrosion distribution of oil and gas pipelines. *Eng. Fail. Anal.* **2016**, *60*, 209–228. [CrossRef]
6.   Bazán, F.A.V.; Beck, A. Stochastic process corrosion growth models for pipeline reliability. *Corros. Sci.* **2013**, *74*, 50–58. [CrossRef]
7.   Qin, H.; Zhang, S.; Zhou, W. Inverse Gaussian process-based corrosion growth modeling and its application in the reliability analysis for energy pipelines. *Front. Struct. Civ. Eng.* **2013**, *7*, 276–287. [CrossRef]
8.   Pan, D.; Liu, J.-B.; Cao, J. Remaining useful life estimation using an inverse Gaussian degradation model. *Neurocomputing* **2016**, *185*, 64–72. [CrossRef]

9.   Peng, W.; Li, Y.-F.; Yang, Y.-J.; Huang, H.-Z.; Zuo, M. Inverse Gaussian process models for degradation analysis: A Bayesian perspective. *Reliab. Eng. Syst. Saf.* **2014**, *130*, 175–189. [CrossRef]

10. Witek, M. Gas transmission pipeline failure probability estimation and defect repairs activities based on in-line inspection data. *Eng. Fail. Anal.* **2016**, *70*, 255–272. [CrossRef]

11. Xie, M.; Tian, Z. Risk-based pipeline re-assessment optimization considering corrosion defects. *Sustain. Cities Soc.* **2018**, *38*, 746–757. [CrossRef]

12. Gong, C.; Zhou, W. First-order reliability method-based system reliability analyses of corroding pipelines considering multiple defects and failure modes. *Struct. Infrastruct. Eng.* **2017**, *13*, 1451–1461. [CrossRef]

13. Benjamin, A.C.; Freire, L.F.; Vieira, R.D.; Cunha, D.J. Interaction of corrosion defects in pipelines e Part 1: Fundamentals. *Int. J. Press. Vessel. Pip.* **2016**, *144*, 56–62. [CrossRef]

14. Benjamin, A.C.; Freire, L.F.; Vieira, R.D.; Cunha, D.J. Interaction of corrosion defects in pipelines e Part 2: MTI JIP database of corroded pipe tests. *Int. J. Press. Vessel. Pip.* **2016**, *145*, 41–59. [CrossRef]

15. Amandi, K.; Diemuodeke, E.; Briggs, T. Model for remaining strength estimation of a corroded pipeline with interacting defects for oil and gas operations. *Cogent Eng.* **2019**, *6*, 1663682. [CrossRef]

16. Sun, J.; Cheng, Y.F. Modelling of mechano-electrochemical interaction of multiple longitudinally aligned corrosion defects on oil/gas pipelines. *Eng. Struct.* **2019**, *190*, 9–19. [CrossRef]

17. Soares, E.; Bruère, V.M.; Afonso, S.M.; Willmersdorf, R.B.; Lyra, P.R.; Bouchonneau, N. Structural integrity analysis of pipelines with interacting corrosion defects by multiphysics modeling. *Eng. Fail. Anal.* **2019**, *97*, 91–102. [CrossRef]

18. Chen, Y.; Zhang, H.; Zhang, J.; Liu, X.; Li, X.; Zhou, J. Failure assessment of X80 pipeline with interacting corrosion defects. *Eng. Fail. Anal.* **2015**, *47*, 67–76. [CrossRef]

19. Kuppusamy, C.S.; Karuppanan, S.; Patil, S.S. Buckling Strength of Corroded Pipelines with Interacting Corrosion Defects: Numerical Analysis. *Int. J. Struct. Stab. Dyn.* **2016**, *16*, 1550063. [CrossRef]

20. Amaro, R.L.; Drexler, E.S.; Slifka, A.J. Fatigue crack growth modeling of pipeline steels in high pressure gaseous hydrogen. *Int. J. Fatigue* **2014**, *62*, 249–257. [CrossRef]

21. Xie, M.; Bott, S.; Sutton, A.; Nemeth, A.; Tian, Z. An Integrated Prognostics Approach for Pipeline Fatigue Crack Growth Prediction Utilizing Inline Inspection Data. *J. Press. Vessel. Technol.* **2018**, *140*, 031702. [CrossRef]

22. Okodi, A.; Lin, M.; Yoosef-Ghodsi, N.; Kainat, M.; Hassanien, S.; Adeeb, S. Crack propagation and burst pressure of longitudinally cracked pipelines using extended finite element method. *Int. J. Press. Vessel. Pip.* **2020**, *184*, 104115. [CrossRef]

23. Zhang, Y.; Xiao, Z.; Luo, J. Fatigue crack growth investigation on offshore pipelines with three-dimensional interacting cracks. *Geosci. Front.* **2018**, *9*, 1689–1697. [CrossRef]

24. Hu, J.; Tian, Y.; Teng, H.; Yu, L.; Zheng, M. The probabilistic life time prediction model of oil pipeline due to local corrosion crack. *Theor. Appl. Fract. Mech.* **2014**, *70*, 10–18. [CrossRef]

25. Lu, B.; Song, F.; Gao, M.; Elboujdaini, M. Crack growth prediction for underground high pressure gas lines exposed to concentrated carbonate–bicarbonate solution with high pH. *Eng. Fract. Mech.* **2011**, *78*, 1452–1465. [CrossRef]

26. Sekhar, A.S. Multiple cracks effects and identification. *Mech. Syst. Signal Process.* **2008**, *22*, 845–878. [CrossRef]

27. Chen, T.; Guestrin, C. *XGBoost: A Scalable Tree Boosting System*; ACM: New York, NY, USA, 2016; Volume 16, pp. 13–17.

28. Nielsen, A.; Mallet-Paret, J.; Griffin, K. Probabilistic Modeling of Crack Threats and the Effects of Mitigation. In Proceedings of the 2014 10th International Pipeline Conference, Calgary, AB, Canada, 29 September–3 October 2014; Volume 3, p. V003T12A021.

29. Sutton, A.; Hubert, Y.; Textor, S.; Haider, S. Allowable Pressure Cycling Limits for Liquid Pipelines. In Proceedings of the 2014 10th International Pipeline Conference, Calgary, AB, Canada, 29 September–3 October 2014; Volume 2, p. V002T06A058.

30. Waheed, B.; McAuliff, K.; Bhuyan, G. Knowledge Gained From a Five-Year Regulatory Compliance Assurance Process for Operators' Pipeline Integrity Management Programs. In Proceedings of the 2016 10th International Pipeline Conference, Calgary, AB, Canada, 26–30 September 2016; Volume 50266, p. V002T01A006.