Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 10, No. 1, 2024

Application of Improved Support Vector Machine in Predicting Failure Pressure of Oil and Gas Pipelines with Internal Corrosion Defects

Wei Yuan, Liang Zhang*, Hao Ren

School of mechanical engineering, Southwest Petroleum University, Chengdu Sichuan 610500, China
*Corresponding author: Liang Zhang

Abstract: In this study, the failure pressure of submarine oil and gas pipelines was predicted by using five commonly used machine learning models and derivative models. Firstly, an efficient local time-intensive finite element algorithm is constructed and used to generate a machine learning database. Secondly, in order to improve the accuracy of machine learning algorithm, the frontier optimization algorithm is improved to form a new predictive regression model IAEO-SVM model. To compare the accuracy of commonly used machine learning models and IAEO-SVM models, a comprehensive evaluation standard consisting of k-fold cross-validation and three statistical indicators was used. Finally, the influence of seawater depth and geometric factors of corrosion defects on the blasting pressure of submarine oil and gas pipelines was comprehensively analyzed through the high-dimensional surface built by the IAEO-SVM model. The IAEO-SVM model shows superior stability and accuracy compared to the comparison model, as demonstrated by its MSE' of 0.0482, R2' of 0.9982, MAE' of 0.1295, and SD of 0.2198. The high-dimensional surface obtained through inversion shows a linear relationship between seawater depth and failure pressure. Meanwhile, the width of corrosion defects has a significant impact on failure pressure, accounting for up to 23% and thus cannot be overlooked.

Keywords: Improved support vector machine, Submarine oil and gas pipelines, Pipeline internal corrosion, Machine learning, Finite element method.

1. Introduction

According to a 2017 report from the American Society of Corrosion Engineers [1], the global economic loss caused by corrosion was estimated to be approximately 2.5 trillion US dollars. In addition, the seabed environment is more susceptible to erosion than land [2], which makes the maintenance and replacement of submarine pipelines more expensive and complex [3]. Therefore, studying the failure pressure of submarine pipelines with corrosion defects is of great significance. It can not only predict the probability of accidents more scientifically and effectively but also provide reliable decision support for pipeline transportation deployment and maintenance detection [4].

At present, a plethora of scholars have conducted extensive research on the failure pressure of corrosion defects. Gao J [5] utilized the finite element method (FEM) to conduct a full-scale blasting test and obtained multiple sets of failure pressure data, suggesting that FEM is a viable approach for predicting the failure pressure of corroded pipelines. Abyani M [6] deployed FEM to assess the reliability of a single corrosion defect in a submarine pipeline. However, this method is intricate to construct the model and consumes a significant amount of time, indicating that it cannot be utilized for large-scale corrosion assessments. Sun M [7] and Chen Y F [8] simplified the corrosion defect morphology into a three-dimensional cuboid shape with only length, width and height, which greatly simplifies the calculation of finite element simulation. ASME B31G-2009 [9], DNVRP-F101 [10] are commonly used methods in the industry for estimating the burst pressure of corroded pipelines. However, these methods only consider the length and depth of corrosion defects, and do not take into account all geometric parameters of defects. As confirmed by Ma B [11], the estimated burst pressure is too conservative, leading to early pipeline replacement and economic losses for enterprises.

Xu W [12] employed ABAQUS software to generate various finite element models and utilized the numerical outcomes to establish a neural network model for predicting the failure pressure of externally corroded pipelines. Nevertheless, the neural network model comprises multiple random initial weight thresholds, which significantly impact the optimization accuracy and efficiency.

In summary, constructing a three-dimensional numerical model of corrosion defects in FEM analysis is a challenging task, and the nonlinear calculation time and cost associated with it are exceedingly high. The outcomes obtained through the traditional formula method are overly conservative, which deviates from the objective of reducing operation and maintenance expenses. The initial random weight threshold of the neural network will bring some instability to the operation results. Currently, numerous studies solely focus on the depth and length of corrosion defects, while neglecting the impact of the width of the corrosion defect and external hydrostatic pressure on pipelines. However, this study reveals that the influence of corrosion width on blasting pressure intensifies with increasing corrosion depth and length. Moreover, the influence of external hydrostatic pressure on the burst pressure cannot be disregarded.

Initially, considering that machine learning needs to build a large number of samples. A novel, efficient PYTHON batch generation model (Chapter 1.2) was proposed and verified via a blasting test conducted by Kang K Y [13]. In this modeling process, we studied the effects of four parameters, namely seawater depth, corrosion defect depth, corrosion defect length, and corrosion defect width, on the failure pressure of

submarine pipelines affected by corrosion defects.

Secondly, the current machine learning algorithms exhibit low accuracy. Therefore, this research proposes an improved machine learning method to replace the traditional prediction technology. The conventional support vector machine (SVM) employs artificially set penalty and loose factors or traditional optimization algorithms, which results in unstable calculation results and low accuracy. The artificial ecological optimization algorithm (AEO) was proposed by Zhao S J [14] as a meta-heuristic algorithm that has shown stronger optimization capabilities and a shorter iteration process when compared to genetic algorithm (GA), particle swarm optimization (PSO), and whale optimization algorithm (WOA). Utilizing AEO to optimize SVM and form the AEO-SVM prediction model can significantly mitigate the instability of the result. Nevertheless, AEO is prone to falling into local optima and exhibits low convergence speed. Hence, the author proposes an improved strategy that incorporates reverse backtracking (LBES) and elite learning to avoid the aforementioned limitations. The resulting enhanced algorithm is called the improved artificial ecological optimization algorithm (IAEO).

Ultimately, the IAEO-SVM prediction model was formulated. Next, 351 sets of finite element analysis results were utilized in seven machine learning models: IAEO-SVM, AEO-SVM, SVM, artificial neural network using Levenberg-Marquardt algorithm (ANN-LM), radial basis function neural network (RBF), extreme learning machine (ELM), and generalized regression neural network (GRNN). The selection of the best prediction model was based on a comprehensive index formed via k-fold cross-validation and three statistical evaluation indexes. Finally, mathematical analysis was conducted to interpret the results.

2. Finite Element Model and Database Construction

2.1. Construction of 3D model

Conventional corrosion defects are irregular, so this paper simplifies them, and the axial projection of corrosion defects is shown in Fig. 1. Then, the internal corrosion defect model was established in ABAQUS by using solid element. The relevant parameters of the three-dimensional solid model are defined as follows:

1) Defect geometric parameters: The pipeline has good symmetry. In order to improve the calculation efficiency, $1/4$ of the pipeline with uniform internal corrosion defects is taken. In Fig. 2, d, l ,and w represent the depth, length and width of the defect respectively.

2) The definition of material and damage: This paper takes API 5L X65 pipeline as an example, which has been widely used in submarine pipeline projects [15]. In the isotropic plastic zone of material nonlinear behavior, NLGEOM in ABAQUS is used to consider geometric nonlinearity. The ultimate condition for the failure of the pipeline is: the defective pipeline is subjected to ultimate internal pressure, and the pipeline material Von-Mises stress reaches the ultimate tensile strength, and the pipeline is destroyed [16]. The remaining mechanical properties and physical properties are shown in Table 1.

3) Material boundary conditions and loads: Fig. 3 shows that both surfaces A and B are subject to displacement constraints from the Z direction; the C and D surfaces are constrained by displacements from the X and Y directions, respectively. The hydrostatic pressure P outside the pipeline is also affected by seawater, which is calculated by formula $P = \rho g h$.

4) Mesh Parameters: The mesh network is constructed utilizing eight-node linear hexahedral simplified integral elements. When compared with tetrahedral elements, their calculation outcomes are more accurate [17]. To account for the quality of the grid, the numerical model is partitioned into dense and sparse areas, as depicted in Fig. 4. The dense area employs a mesh size of 0.5-1.2mm. A blasting test was conducted on X65 pipeline by Kang, and the experimental results were compared with the simulation outcomes, as shown in Table 2.

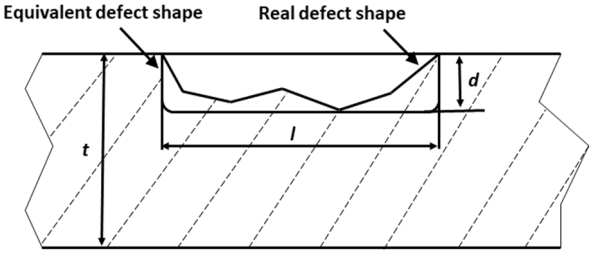

Figure 1. Simplified schematic diagram of corrosion defects.

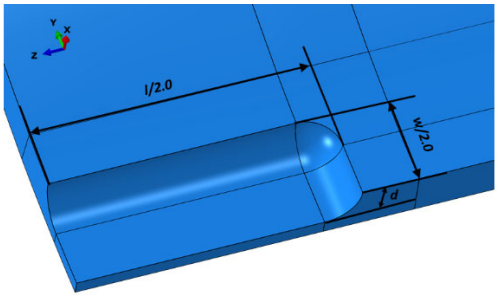

Figure 2. Geometric parameters of corrosion defects.

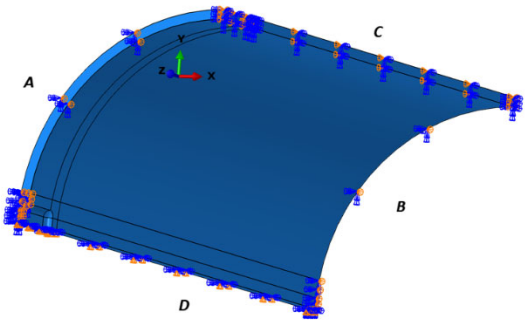

Figure 3. Boundary condition diagram.

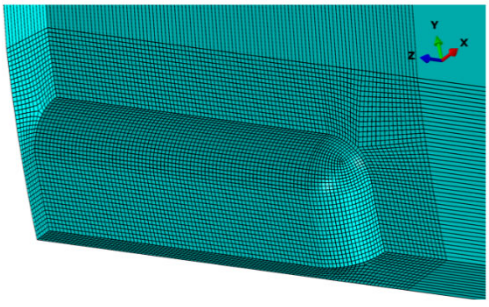

Figure 4. Grid division diagram

Table 1. Mechanical and physical properties of corrosion defects.

Parameter	Symbol	Value	Parameter	Symbol	Value
Pipeline outer diameter	D_o	762mm	Steel Poisson's ratio	ν	0.3
Young's modulus	E	2.1GPa	Steel density	ρ_S	7850 kg/m^3
Engineering yield stress	S_{EY}	464.5MPa	Sea water density	ρ_W	1020kg/m^3
Engineering ultimate stress	S_{EU}	563.8MPa			

Table 2. Simulation results verification

number	Corrosion width w (°)	Corrosion length l (mm)	Corrosion depth d (mm)	Pipeline outer diameter D_o (mm)	Pipeline wall thickness t (mm)	Experiment (MPa)	Finite element simulation (MPa)	Relative error (%)
NO.1	10	200	4.4	762	17.5	24.11	24.90	0.03%
NO.2	10	200	8.8	762	17.5	21.74	21.30	2.02%
NO.3	10	100	8.8	762	17.5	24.30	24.30	0
NO.4	10	300	8.8	762	17.5	19.08	18.70	-1.99%

2.2. Database construction of machine learning prediction model

Based on the aforementioned analysis, it is evident that the computational expense of FEM is substantial. Therefore, the SVM model, which exhibits exceptional nonlinear fitting capabilities, is an optimal choice for enhancing computational efficiency [18]. To generate the internal corrosion defect model in batches, the methodology outlined in Chapter 1.1 is implemented using a Python script. It is worth noting that the script construction steps are:

1) Firstly, the model is established in accordance with Chapter 1.1, and the PYTHON script for the generated model is parameterized.

2) The model is set to comprise of two load steps. The first load step is designed to apply the seawater pressure, with a total loading time of 1 second. The second load step is used to apply pressure inside the tube, with a total loading time of 1 second.

3) To achieve a balance between computational efficiency and accuracy, the internal pressure was set at a fixed value of 100 MPa, and a linear loading approach was adopted. Given that the pipeline cannot withstand the internal pressure, an local intensive time step was employed, which enabled the pipeline to be destroyed within the local intensive time interval. The local intensive time interval was determined using Eq. (1). The time interval of the local intensive interval was set to 0.001, which allowed for the calculation result to attain an accuracy of 0.1 MPa.

$$\text{time\_start} = (1\text{-}20\%) \bullet \frac{2t\sigma_u}{(D-t)}\left[\frac{1-d/t}{1-d/tM}\right]/100 + S\_t/10000 \quad (1)$$

Where: time_start is the starting point of time, where $\frac{2t\sigma_u}{(D-t)}\left[\frac{1-d/t}{1-d/tM}\right]$ is the pressure value obtained by the DNV-RF-F01 formula method, $(1\text{-}20\%)$ is the 80 % pressure value of the formula method, and the formula method is divided by 100 to convert the time unit to 0-1s. $M = \sqrt{1+0.31\frac{l^2}{Dt}}$. The same is true for the depth of the ocean. $S\_t/10000$ is the time node effect of ocean depth estimation. The end node of local intensive time is uniformly set to $\text{time\_end} = 0.40$.

4) Upon reaching a certain period of time, the limit tensile stress is assessed to determine whether it has been reached. If this is the case, proceed to Step 1 and solve the next INP file. On the other hand, if the limit tensile stress has not been reached, return to the Step 3 solution.

5) Retrieve the frame number for the maximum stress from the ODB file, obtain the corresponding time for this frame number, and decode this time by multiplying it by 100 to obtain the final burst pressure of the pipeline.

The outer diameter D_o was 762 mm. The pipeline length is $3\times D_o/2$; the length will not affect the analysis [21]. The thickness t of X65 pipeline is 17.5 mm. It should be noted that the model is highly symmetrical, so only 1/4 of the corrosion defects can be analyzed. Numerical calculations for fluid pressure are performed on 351 sets of diverse inputs, as provided in Table 3.

The range of defect length and depth was mainly referenced from previous studies [22]. However, the starting point for defect length in this study is 150 mm, which may not accurately reflect the influence of corrosion parameters and seawater depth on blasting pressure. In this paper, through multiple orthogonal experiments, we present the ideal ranges of influence caused by each defect value on blasting pressure in Table 3. It should be noted that the machine learning database is limited to this range, which also characterizes the limitations of the application model in this study.

Table 3. Corrosion shape parameters

	Defect length l (mm)	Defect depth d(mm)	Corrosion width a (mm)	Sea depth S_t (m)
Value ranges	60 : 20 : 300	0.3t : 0.2t : 0.7t	10 : 10 : 30	100, 250, 500

3. Support Vector Machine and The Theoretical Basis of IAEO Algorithm

3.1. Theoretical basis of support vector machine

Support vector machine is proposed by Sain S [23], which regresses and classifies data through supervised learning. Suppose there exists a sample data set $\{(x_1,y_1),(x_2,y_2),...(x_n,y_n)\}$, where $x_i \in R^n$ and n are the number of samples, and the data set is linearly separable. The discriminant function can be expressed as:

$$f(x) = w \bullet x + b \tag{2}$$

Where: w is the inertia weight; b is the intercept of hyperplane.

The classification hyperplane expression corresponding to the discriminant function is Eq. (3).

$$w \bullet x + b = 0 \tag{3}$$

Normalize $f(x)$ so that any sample satisfies $|f(x)| \geq 1$, then the sample $|f(x)| = 1$ closest to the decision surface satisfies the condition:

$$y_i(w \bullet x + b) - 1 \geq 0, \quad i = 1, 2, 3, \ldots, n \tag{4}$$

The interval surface distance is: $2/\|w\|$. In order to maximize the interval distance, it is necessary to find the minimum value of $\|w\|$. For the linearly inseparable case, the slack variable ξ_i can be introduced to achieve the purpose of accurately dividing the training samples, which can be expressed as the optimization problem shown in Eq. (5):

$$\begin{aligned} f &= \min_{w,\ b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\xi_i \\ \text{s.t. } & y_i(w \bullet x + b) \geq 1 - \xi_i \end{aligned} \tag{5}$$

Where: $\xi_i \geq 0$, C are penalty factors, $C>0$.

For nonlinear problems, the kernel function is introduced:

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \tag{6}$$

Where : ϕ is the mapping from the original space to the feature space, Fig. 5 is the schematic diagram of the mapping space.

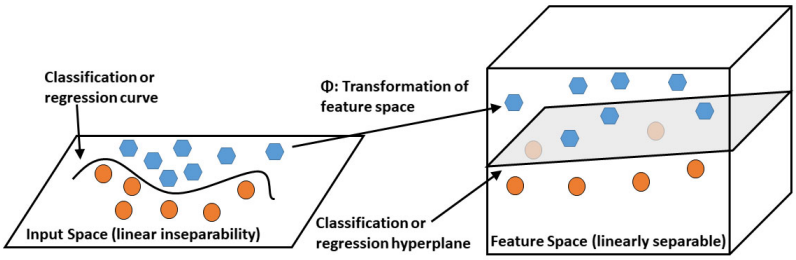

Figure 5. Schematic diagram of transformation from original space to feature space

The objective function of nonlinear regression SVM is:

$$\begin{aligned} f &= \max \sum_{i=0}^{l}\alpha_i - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\ \alpha_j\ y_i\ y_j K(x_i, x_j) \\ & \sum_{i=0}^{l}\alpha_i\ y_i = 0 \end{aligned} \tag{7}$$

Where: $0 \leq \alpha \leq C$; i=1, 2, ..., l.

In solving practical problems, RBF is widely used in SVM regression because of its excellent kernel function performance. Its mathematical expression is as follows:

$$K(x_i, x_j) = \exp\left[-\frac{|x_i - x_j|^2}{\sigma^2}\right] \tag{8}$$

Where: σ is the radius of the radial basis, let $\gamma = 1/\sigma^2$, into Eq. (8):

$$K(x_i, x_j) = \exp\left[-\gamma |x_i - x_j|^2\right] \tag{9}$$

According to the Eqs. (5-9), the core point of the SVM algorithm is to find the appropriate penalty factor C and the appropriate radial basis kernel function parameters γ.

3.2. AEO algorithm

As demonstrated in Chapter 2.1, the prediction accuracy of SVM is influenced by the penalty factor and the parameters of the radial basis kernel function. Research has shown that optimization algorithms can considerably enhance the accuracy of SVM.

The AEO algorithm, a new meta-heuristic algorithm proposed by Zhao W [24], is inspired by the energy flow in ecosystems and the behavior of producers, consumers, and decomposers. It primarily consists of the production stage, consumption stage, and decomposition stage. The organisms in each stage represent feasible solutions to the optimization problem. Through the process of energy flow, the theoretical optimal solution is gradually approached through iteration.

1) Production stage: The producer x_1 (the worst solution in the iterative process) is iteratively updated by the linear coupling of the decomposer x_n (the best individual in the iterative process) and the random individual x_{rand}, which is used to guide the evolution direction of the consumer. Its expression is:

$$x_1(t+1)=(1-\alpha)x_n(t)+\alpha x_{\text{rand}}(t) \tag{10}$$

Where: n is the population size, $\alpha=(1-t/T)\cdot r_1$, where r_1 is the weighted coefficient, and $r_1\in[0,1]$. t and T are the current iteration number and the maximum iteration number. x_{rand} is the vector of random position in the boundary interval.

2) Consumption stage: consumers are herbivores, carnivores and omnivores, which are divided according to a certain proportion in the model. And iterative evolution in three different postures.

Case1: Herbivores x_i are updated only by producers x_1, and the mathematical expression of their behavior is:

$$x_i(t+1)=x_i(t)+C\cdot(x_i(t)-x_1(t)) \tag{11}$$

Where: $i\in[2,n]$, $C=v_1/2|v_2|$,which v_1 and v_2 obey normal distribution, C is the improved multiplication operator of levy flight.

Case 2: Carnivores x_i randomly prey on animals x_j with higher energy levels than themselves. The mathematical expression of their behavior is:

$$x_i(t+1)=x_i(t)+C\cdot(x_i(t)-x_j(t)) \tag{12}$$

Where: $i\in[3,n]$, $j\in\text{randi}([2,i-1])$.

Case3: Omnivores x_i can prey on producers x_1 and an animal x_j with a higher energy level than itself. The mathematical expression of its behavior is:

$$x_i(t+1)=x_i(t)+C\cdot(r_2\cdot(x_i(t)-x_1(t))+(1-r_2)(x_i(t)-x_j(t))) \tag{13}$$

Where: $j\in\text{randi}([2,i-1])$, $r_2\in[0,1]$ is herbivorous proportion.

3) Decomposition stage: The organisms in the ecosystem will eventually be decomposed by the decomposer x_n, and the mathematical expression of their behavior is:

$$x_i(t+1)=x_n(t)+B\cdot(e\cdot x_n(t)-h\cdot x_i(t)) \tag{14}$$

Where: $i\in[1,n]$, $B=3u$ is the decomposition factor, which u obeys the standard normal distribution. $e=r_3\cdot\text{randi}(|1,2|)-1$, $h=2r_3-1$, $r_3\in[0,1]$ is the weight factor.

After the above iteration, if the optimal solution does not meet the accuracy condition or the maximum number of iterations, the process is repeated until the conditions are met. However, in terms of optimization strategy, the linear decrease in the producer's convergence factor is insufficient to provide a wide search range in the early stage of the search, which limits the algorithm's accuracy. Therefore, we propose two improvement strategies to address the drawbacks of linear convergence.

3.3. Improvement Strategy

3.3.1. Local backtracking mining strategy (LBES)

It has been confirmed by Zhao that the AEO algorithm can further improve local development capabilities. The decomposer is the optimal solution after each iteration, and its historical sequence contains important optimization information. The local backtracking exploitation strategy (LBES) can be used to achieve re-mining in the local area through inheritance learning, as shown in Fig. 6.

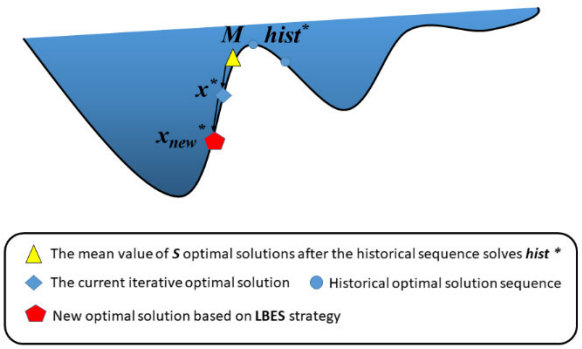

Figure 6. The schematic diagram of LBES strategy with backtracking step of three.

The darker the blue in the figure, the closer it is to the real global optimal value. When the length N of the historical optimal solution sequence $hisx^*$ exceeds the local backtracking step s, the LBES strategy will be excited and reversely select the backtracking step s in $hisx^*$. The optimal solution sequence $hisx^*$, calculate its mean value M and act on the current decomposer x^* to generate a new decomposer x_{new}^*. The corresponding expression is:

$$\begin{cases} x_{new}^{*} = x^{*} + rand \cdot \left(M - x^{*}\right) \\ M = \overline{hisx}^{*}\left(N - s + 1 : N\right) \end{cases} \tag{15}$$

Where: $rand \in [0, 1]$, whether the new decomposer x_{new}^{*} is retained will be determined by Eq. (16) :

$$x^{*} = \begin{cases} x_{new}^{*}, \ f\left(x_{new}^{*}\right) < f\left(x^{*}\right) \\ x^{*}, otherwise \end{cases} \tag{16}$$

It is noteworthy, the larger the value of the backtracking step s is, the more historical information is used, the more noise is contained, and the higher the complexity.

3.3.2. Elite reverse learning strategy

Tizhoosh H R [25] proposed a reverse solution that is closer to the global optimal. The main update principle is to sort the excellent individual n and the excellent individual reverse solution. The individuals with lower fitness are selected as the pre-iteration individuals. The definition is:

$$x_{new\_elite}^{*} = rand \cdot (\alpha + \beta) - x_{old\_elite}^{*} \tag{17}$$

Where: $x_{new\_elite}^{*} \in [\alpha, \beta]$, α, β are the upper and lower boundaries.

The dynamic boundary overcomes the disadvantage of fixed boundary's difficulty in preserving search experience, enabling it to search in narrower spaces and improve convergence speed. However, the fixed boundary can avoid getting trapped in local optima during the optimization process. As LBES enhances local optimization capability in this paper, we choose the fixed boundary strategy to strengthen our ability to escape local optima. The optimization principle is illustrated in Fig. 7. Although the dynamic boundary improves convergence speed, it may lead to getting trapped in local optima.

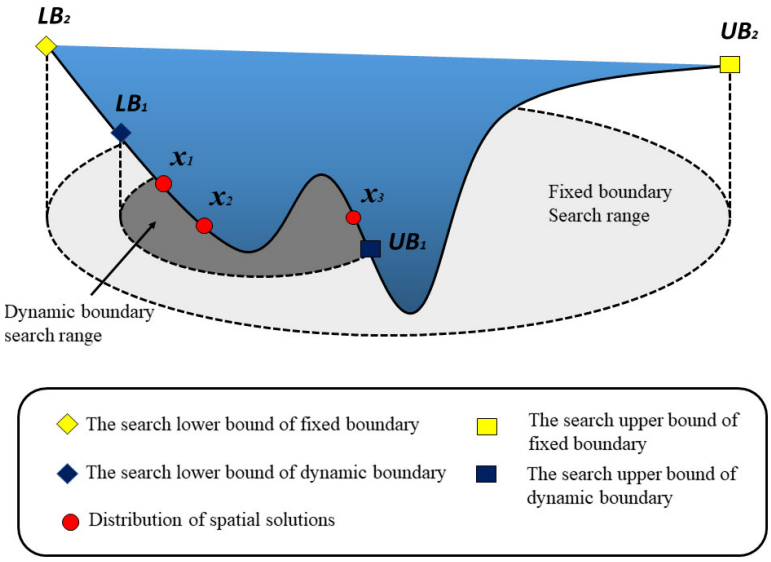

Figure 7. Elite reverse learning dynamic boundary and fixed boundary diagram

3.4. IAEO-SVM model construction and experimental process

To avoid the influence of data on each other, it is necessary to normalize all the data due to their varying dimensions.

$$y = 2\frac{x - x_{min}}{x_{max} - x_{min}} - 1 \tag{18}$$

$$y_r = \frac{1}{2}\left(x_{sim} + 1\right)\left(x_{max} - x_{min}\right) + x_{min} \tag{19}$$

Where: x represents the parameter to be normalized, while x_{min} and x_{max} respectively denote the maximum and minimum values of the variable. y represents the normalized value of the variable, x_{sim} denotes the fitted output value, and y_r represents the model prediction value obtained through reverse normalization.

From the SVM theoretical knowledge presented in Section 2.1, it is apparent that the penalty factor and the slack variable of the SVM model are manually set or optimized using an optimization algorithm. Currently, the optimization algorithms commonly employed fail to satisfy the prediction accuracy requirements of SVM. Consequently, this paper proposes an improved optimization algorithm, the IAEO algorithm, to optimize SVM. The flow chart of the IAEO-SVM optimization model is illustrated in Fig. 8. It is worth noting that in order to ensure the fairness of the comparison between the IAEO-SVM model and the AEO-SVM model, this paper sets the population number of the optimization algorithm to 10, iterates 20 times, and the optimization boundary is set to [0.1, 100].

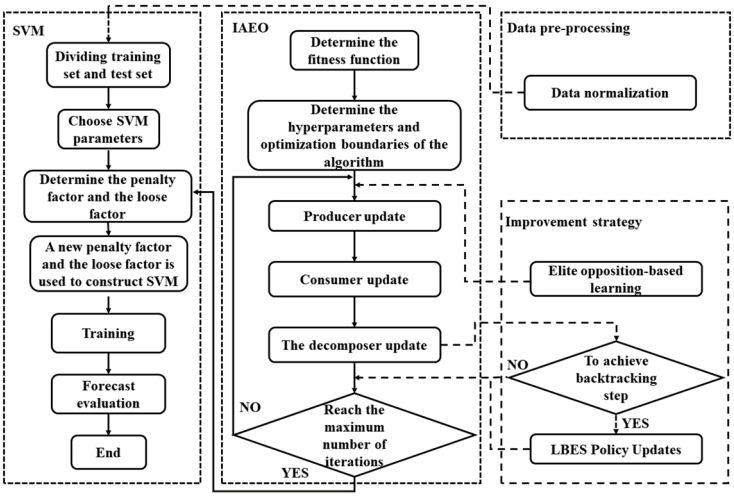

Figure 8. IAEO-SVM model flow chart

4. Key Parameters Setting and Evaluation Index of Machine Learning Model

4.1. Key parameter settings of machine learning model

The parameter settings of the machine learning model can greatly affect the final results [26]. After repeated testing, the final model's parameters are determined as shown in Table 4.

It is noteworthy that the ANN training will not use the traditional gradient descent algorithm, as this algorithm can be easily influenced by the initial values [28], which can lead to premature convergence in regions with gentle gradients. The LM algorithm, proposed by Wilamowski B M [29], can significantly improve convergence speed and accuracy compared to gradient descent. Therefore, in this paper, we use the LM algorithm instead of the ANN gradient descent algorithm. ELM is widely used in the fields of image recognition and fault diagnosis [30]. RBF and GRNN have excellent performance in nonlinear fitting in engineering field.

Table 4. Machine learning model related parameter settings.

Model	Parameter
ANN-LM	Activation function from input layer to hidden layer: logsig Hidden layer to output layer activation function: purlin Number of hidden layer neurons: 8 Learning goals: 1e-8 Learning rate: 0.1 Maximum number of iterations: 1000 Training function: trainlm
SVM	Kernel function: radial basis c: 2.1 g: 2.1
RBF	The extension speed of radial basis function: 1.5 Hidden layer neurons: 8
GRNN	Smooth factor value : 1.0
ELM	Activation function from input layer to hidden layer: sig Number of hidden layer neurons: 8

4.2. Evaluation indicators

Verify the pros and cons of the model usually Means Absolute Error (MAE), Means Square Error (MSE), and Coefficient of Determination (R^2) are three indicators to measure. Evaluating indicators can effectively reflect the accuracy of model prediction and fitting degree, the Eqs. (20-22) is:

$$E_{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - p_i\right| \tag{20}$$

$$E_{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - p_i\right)^2 \tag{21}$$

$$R^2 = \frac{\sum_{i=1}^{n}\left(p_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{22}$$

Where p_i is the prediction result, y_i is the sample label, $\bar{y}$ Is the sample mean value, and n is the number of

samples.

The hold-on verification method has a strong correlation with data grouping, resulting in a better regression effect. However, to avoid incomplete utilization of data information, this paper uses the K-fold cross-validation method to calculate the mean value of the evaluation index. Therefore, the evaluation formula presented above is derived as Equation (23-25).

$$E'_{MAE} = \frac{1}{k}\sum_{j=1}^{k}\frac{1}{n}\sum_{i=1}^{n}\left|y_i - p_i\right| \tag{23}$$

$$E'_{MSE} = \frac{1}{k}\sum_{j=1}^{k}\frac{1}{n}\sum_{i=1}^{n}\left(y_i - p_i\right)^2 \tag{24}$$

$$R^{2\prime} = \frac{1}{k}\sum_{j=1}^{k}\frac{\sum_{i=1}^{n}\left(p_i - \overline{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \tag{25}$$

Where j is the K-fold of j-fold cross-validation.

5. Results and Numerical Analysis

5.1. Comparison of various machine learning methods

The k value in k-fold cross-validation is set to 9 to ensure a more uniform distribution of data between the test and training sets. The comprehensive evaluation index in Chapter 3.2 is utilized to assess the performance of the seven machine learning models. A smaller value of comprehensive indicators E'_{MAE} and E'_{MSE} indicates better performance, while a value of $R^{2\prime}$ closer to 1 is also desirable. Notably, the $R^{2\prime}$ value of SVM, IAEO-SVM, AEO-SVM, and ANN-LM in the seven machine learning algorithms exceeds 0.99. Moreover, based on the E'_{MAE} and E'_{MSE} indicators, these models outperform ELM, RBF, and GRNN. Consequently, the following analysis is focused on the four superior models.

The training set (see Table 5) E'_{MAE} and E'_{MSE} values of the ANN-LM model were higher than those of the SVM model and its derivative models (AEO-SVM, IAEO-SVM). This finding suggests that the ANN-LM model may not have fully learned the experiential knowledge embedded in the training data, resulting in a lower training efficacy compared to SVM and its derivatives. The performance of ANN-LM in the test set is similar to that of SVM, but much lower than that of SVM derivative model, which confirms its under-fitting in the training stage. Table 5 presents the outcomes of the training set, and the optimal machine learning model is SVM. However, it is noteworthy that the penalty factor and relaxation factor of SVM are artificially designated, thereby resulting in an over-fitting state that undermines the accuracy of machine learning. By comparing the results of the training set and test set of SVM, we observe a decline in the average coefficient of determination $R^{2\prime}$ from 0.9997 to 0.9961.

The AEO-SVM model mitigates the occurrence of the over-fitting state to a certain extent, leading to an increase in the $R^{2\prime}$ value of the test set to 0.9980. Nevertheless, this improvement comes at the cost of an increase in the E'_{MAE} and E'_{MSE} values of the training set and a decrease in the $R^{2\prime}$ value, resulting in an under-fitting state. As Chapter 2 illustrates, this issue stems from the AEO algorithm falling into a local optimum. Conversely, the comprehensive evaluation index E'_{MAE} and E'_{MSE} values of the IAEO-SVM model in the training set are lower than AEO-SVM, while the $R^{2\prime}$ value of IAEO-SVM model is better than AEO-SVM. These results suggest that the IAEO-SVM model identifies better support vectors during the training process and acquires more experiential knowledge.

Table 6 displays the outcomes of the test set, which depict lower comprehensive evaluation indicator E'_{MAE} and E'_{MSE} values for the IAEO-SVM model relative to AEO-SVM, while the $R^{2\prime}$ value of IAEO-SVM model exceeds that of AEO-SVM. These findings indicate that the IAEO-SVM model exhibits superior robustness compared to AEO-SVM, and can effectively mitigate the over-fitting caused by the artificial setting of the SVM model and the under-fitting resulting from the local optimum of the AEO-SVM model.

Table 5. K-fold cross validation of training set evaluation index

Training	ANN-LM	ELM	RBF	GRNN	SVM	AEO-SVM	IAEO-SVM
E'_{MAE}	0.1727	0.0837	1.71554	2.2922	1.2993	0.0837	0.0835
E'_{MSE}	0.0641	3.0193	4.2836	7.7402	0.0088	0.0138	0.0135
$R^{2\prime}$	0.9978	0.9005	0.8584	0.7442	0.9997	0.9995	0.9996

Table 6. K-fold cross validation test set evaluation index

Testing	ANN-LM	ELM	RBF	GRNN	SVM	AEO-SVM	IAEO-SVM
E'_{MAE}	0.2088	0.1886	1.7408	2.3811	1.3378	0.1343	**0.1295**
E'_{MSE}	0.1010	3.2047	4.4492	8.4077	0.1043	0.0520	**0.0482**
$R^{2\prime}$	0.9961	0.8823	0.8389	0.7034	0.9961	0.9980	**0.9982**

Based on statistical indicators, it is evident that among the aforementioned machine learning models, the IAEO-SVM, SVM, and ANN-LM models exhibit superior performance. To gain a more intuitive understanding of the error distribution of these three models, the true values of the nine test sets are subjected to linear regression analysis with the predicted

values of the models. Notably, the application of the IAEO algorithm has optimized the goodness of fit R^2 of the SVM model, increasing it from 0.9969 to 0.9984 (see Fig. 10(a), (c)), thereby enhancing its accuracy.

The residual analysis of test set 1 reveals that the errors of the three models conform to the normal distribution. Specifically, the IAEO-SVM model exhibits a mean residual error of -0.0081 MPa in Fig. 10 (a), which is an order of magnitude greater than the accuracy of the SVM model. Analysis of variance (SD) indicates that only the variance of the IAEO-SVM model is less than 0.3 in Fig. 11 (a), while the other models exceed 0.3, underscoring the superior robustness of the IAEO-SVM model.

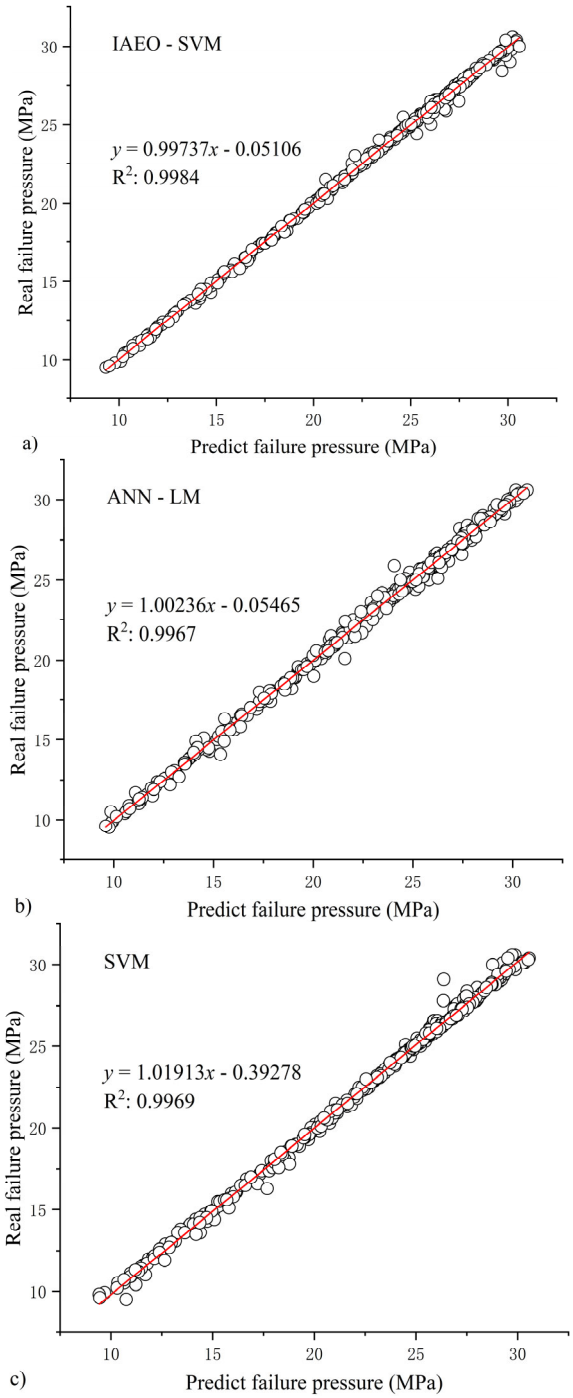

Figure 9. Goodness of fit of IAEO-SVM (a), ANN-LM (b) and SVM (c) models.

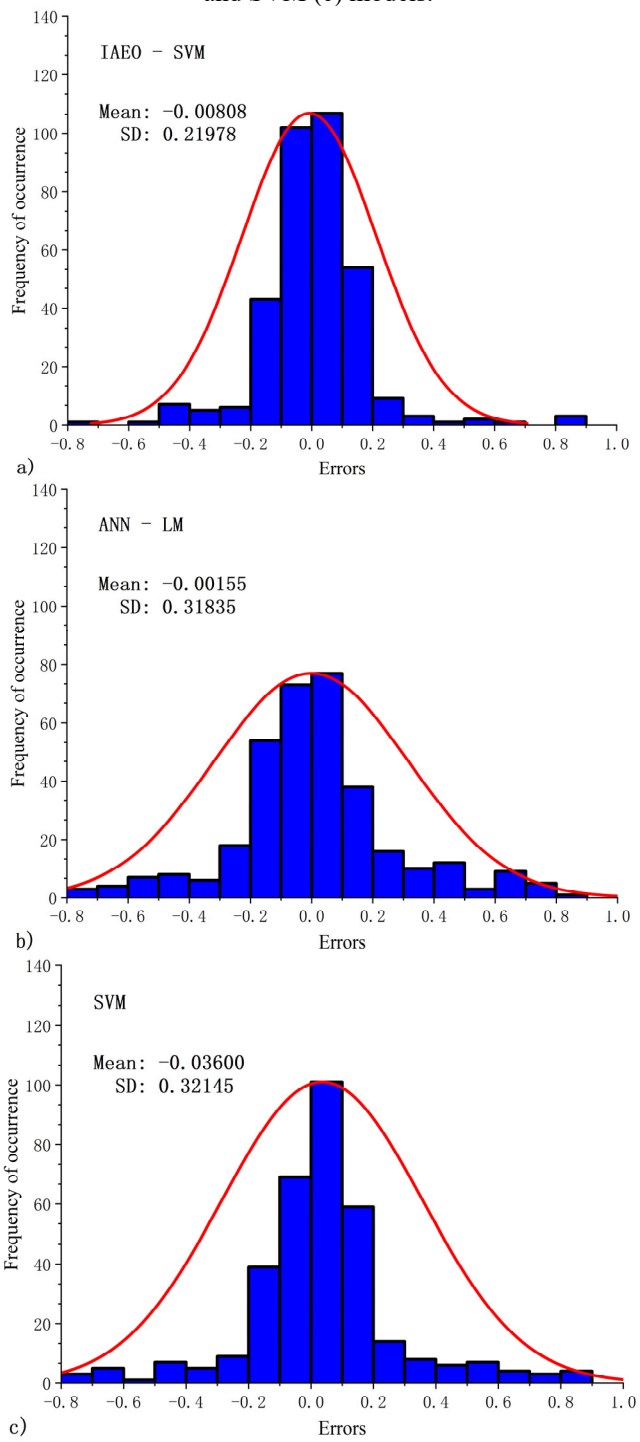

Figure 10. Residual and variance plots of IAEO-SVM (a), ANN-LM (b) and SVM (c) models.

5.2. Mathematical analysis of IAEO-SVM fitting high-dimensional surface

The input data is inwardly interpolated based on the poles used to fit the geometric parameters of corrosion defects, with the number of interpolations set at 100. Subsequently, the interpolated data is fed into the IAEO-SVM model, resulting in high-dimensional spatial surfaces with seawater depths of 100 m, 250 m and 500 m (see Fig 11).

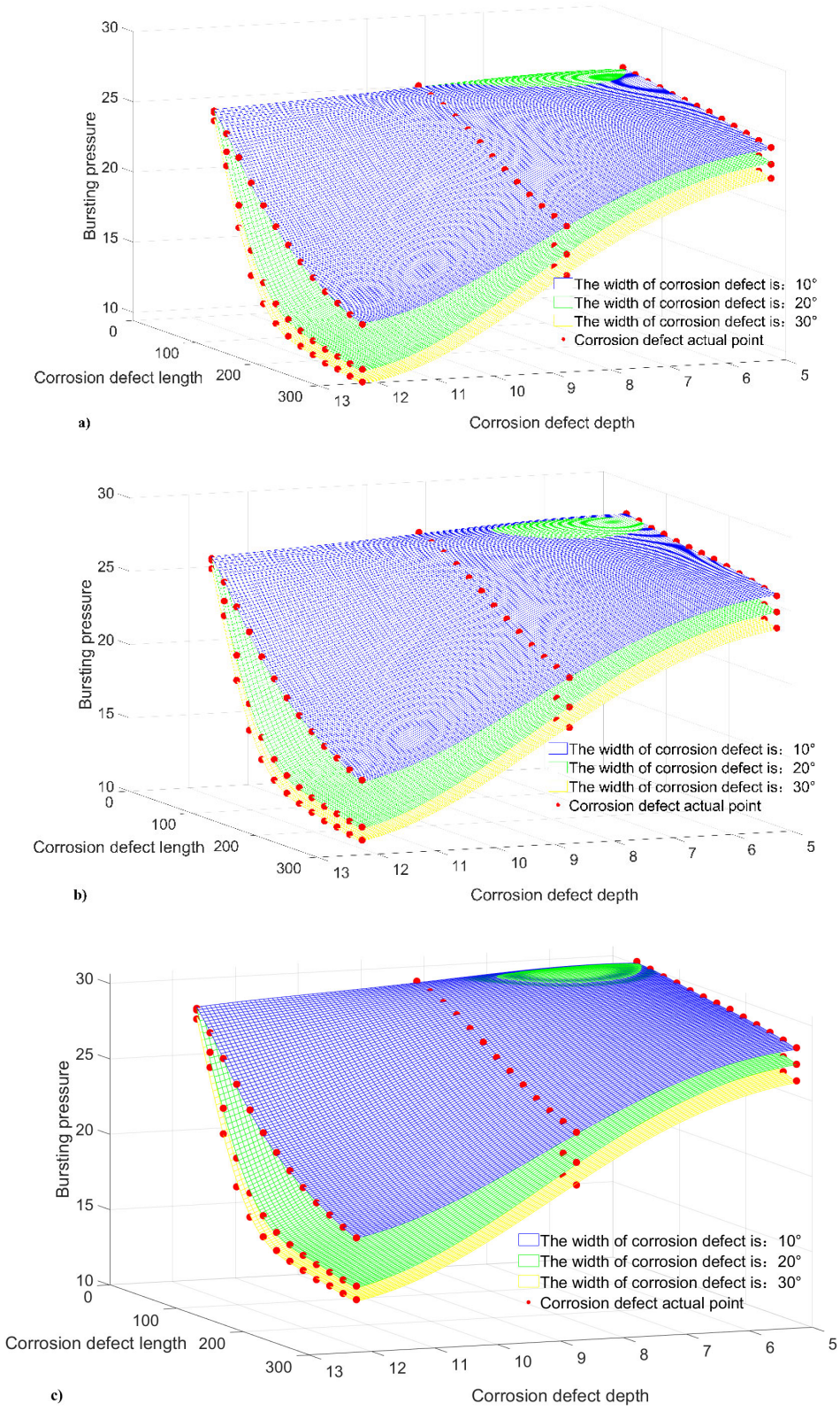

Figure 11. The effect of seawater depth and corrosion defects on blasting pressure. a)100m, b) 250m, c)500m

Upon observing Fig. 11, it is evident that blasting pressure exhibits a linear increase as ocean depth increases. This indicates that greater ocean depths can aid in safeguarding defective pipelines to a certain extent.

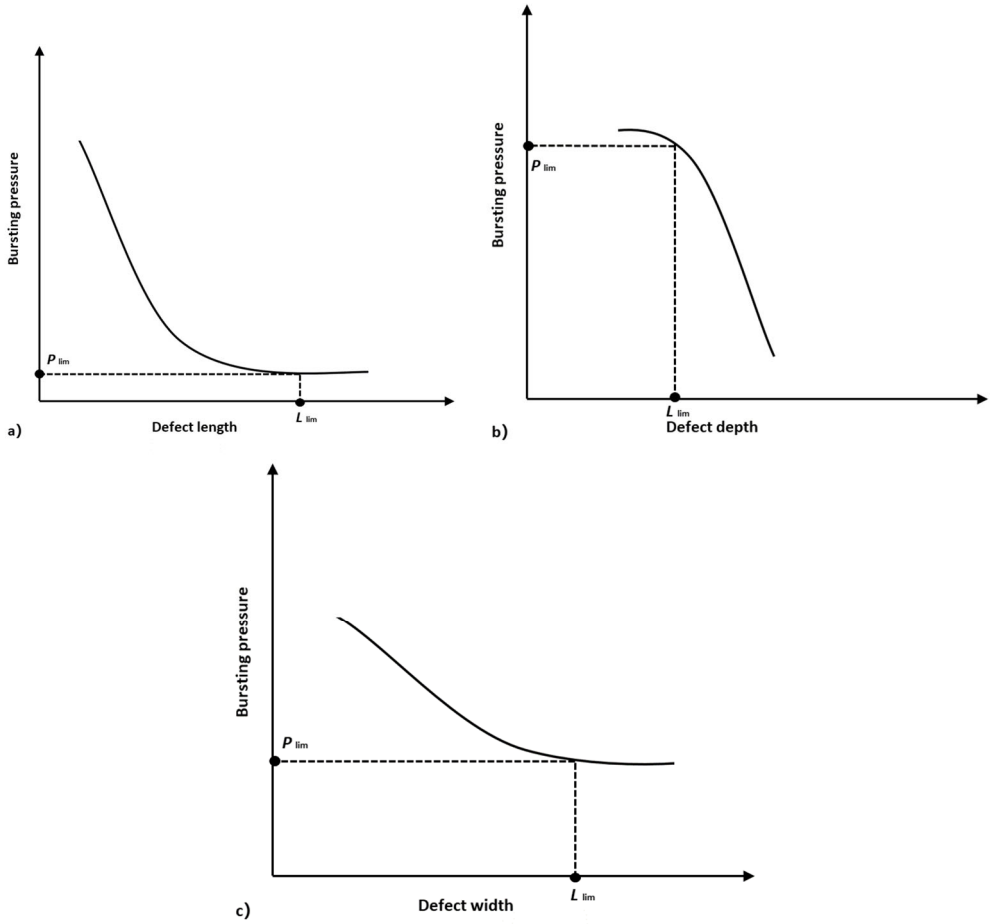

Figure 12. Trend diagram of the influence of length(a), depth(b) and width(c) of defects on blasting pressure.

Upon examining Fig 12, it is evident that blasting pressure exhibits the characteristics of a downward-opening quadratic function with an increase in defect depth. The steepness of this feature demonstrates significant positive correlation with corrosion width and length. This indicates that an increase in corrosion width and length can accelerate the decrease of blasting pressure. Similarly, an increase in corrosion defect length causes blasting pressure to exhibit the characteristics of an upward-opening quadratic function. The steepness of this feature also bears significant positive correlation with the depth and width of the corrosion defect. The upward-opening suggests the existence of a limit threshold, beyond which the burst pressure of the pipeline remains unaffected.

On this basis, the influence of corrosion width is observed. With the increase of width, the reduction rate of blasting pressure increases first and then decreases. It shows that the influence of corrosion defect width on blasting pressure will increase with the increase of width, and the influence will gradually disappear after increasing to a certain value. The change trend is similar to the influence of corrosion length on blasting pressure. There is a limit threshold for the width of the corrosion defect after 30 °. After exceeding this threshold, it can be approximated as having no effect on the burst pressure of the pipeline. When the length of the corrosion defect is fixed at 200 mm and the depth is 0.7 times the wall thickness, the width of the corrosion defect changes from 10 ° to 30 °, which will cause the relative change of the failure pressure to exceed 23 %. It shows that the influence of corrosion width on blasting pressure cannot be ignored when the defect length is long and the defect depth is deep. This finding can also be seen from Fig.11 (b, c).

6. Conclusion

The conclusion is summarized as the following four points:

1) The IAEO algorithm is obtained by improving the AEO algorithm with LBES strategy and reverse elite strategy. LBES improves the accuracy of the algorithm in the later stage, and the reverse elite strategy improves the diversity of the population. The two algorithms are applied to the SVM model respectively. Experiments show that the calculation accuracy, stability and robustness of the IAEO algorithm are improved.

2) The average goodness of fit of RBF, GRNN and ELM models is less than 0.9, and the average MSE is greater than 3.0, indicating that these models are not suitable for the failure pressure prediction of submarine oil and gas pipelines. Compared with the SVM model, the IAEO-SVM model can better avoid the occurrence of 'over-fitting'. Compared with the AEO-SVM model, the determination coefficient and relative error of both the training set and the test set are greatly improved. Compared with the ANN-LM model, the IAEO-SVM model has smaller variances, indicating that IAEO-SVM has better robustness. Considering the IAEO-SVM is a kind of high precision, more stable submarine oil and gas pipeline failure pressure prediction model.

3) With the increase of width, the reduction rate of blasting pressure increases first and then decreases. It shows that the influence of corrosion defect width on blasting pressure will increase with the increase of width, and the influence will gradually disappear after increasing to a certain value.

4) The existing empirical methods advocate neglecting the width of corrosion defects, but through the research

conducted in this paper, it has been found that: When the depth of seawater, the length and depth of corrosion defects are fixed, and the width of corrosion defects changes from 10 ° to 30 °, the maximum relative change of blasting pressure will exceed 23 %. It shows that the influence of corrosion width on blasting pressure cannot be ignored.

7. Conflict of Interest

There are no known competing economic interests or personal relationships.

Acknowledgment

Thank you for the experimental environment provided by the key laboratory of the Ministry of Education of Southwest Petroleum University, as well as the hard work of the editing teacher and the review teacher.

References

[1] G. Teran, S. Capula-Colindres, J.C. Velazquez, M.J. Fernandez-Cueto, D. Angeles-Herrera, H. Herrera-Hernandez, Failure Pressure Estimations for Pipes with Combined Corrosion Defects on the External Surface: A Comparative Study, International Journal of Electrochemical Science, (2017) 10152-10176.

[2] C. Bao-ping, Z. Yan-ping, Y. Xiao-bing, G. Chun-tan, L. Yong-hong, C. Guo-ming, L. Zeng-kai, J.I. Ren-jie, A Dynamic-Bayesian-Networks-Based Resilience Assessment Approach of Structure Systems: Subsea Oil and Gas Pipelines as A Case Study, CHINA OCEAN ENGINEERING, (2020) 597-607.

[3] M. Pourahmadi, M. Saybani, Reliability analysis with corrosion defects in submarine pipeline case study: Oil pipeline in Ab-khark island., OCEAN ENGINEERING, (2022).

[4] L. Zhang, Y. Gao, Improvement of BP neural network model and its application in predicting external corrosion rate of submarine pipeline, Journal of Safety and Environment, (2022). https://doi.org/10.13637/j.issn.1009-6094.2022.1608.

[5] J.A. Gao, P.A. Yang, X.A.L.D. Li, J.A. Zhou, J.A. Liu, Analytical prediction of failure pressure for pipeline with long corrosion defect., OCEAN ENGINEERING, (2019) 106497.

[6] M. Abyani, M.R. Bahaari, A new approach for finite element based reliability evaluation of offshore corroded pipelines., International Journal of Pressure Vessels & Piping, (2021) 104449.

[7] M. Sun, H. Zhao, X. Li, J. Liu, Z. Xu, A new evaluation method for burst pressure of pipeline with colonies of circumferentially aligned defects., OCEAN ENGINEERING, (2021) 108628.

[8] Y. Chen, F. Hou, S. Dong, L. Guo, T. Xia, G. He, Reliability evaluation of corroded pipeline under combined loadings based on back propagation neural network method, OCEAN ENGINEERING, (2022) 111910.

[9] B. Ma, J. Shuai, J. Wang, K. Han, Analysis on the Latest Assessment Criteria of ASME B31G-2009 for the Remaining Strength of Corroded Pipelines, Journal of Failure Analysis and Prevention, (2011) 666-671.

[10] R. Zhou, X. Gu, X. Luo, Residual strength prediction of X80 steel pipelines containing group corrosion defects, OCEAN ENGINEERING, (2023) 114077.

[11] B.M.B. Ma, J.S.J. Shuai, D.L.D. Liu, K.X.K. Xu, Assessment on failure pressure of high strength pipeline with corrosion defects, ENGINEERING FAILURE ANALYSIS, (2013) 209-219.

[12] W. Xu, C.B. Li, J. Choung, J. Lee, Corroded pipeline failure analysis using artificial neural network scheme, ADVANCES IN ENGINEERING SOFTWARE, (2017) 255-266.

[13] K.Y. Kang, 2022. Research on Residual Strength of Corrosion Defect Pressure Piping Based on ANSYS. Shenyang University of Chemical Technology. https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD-202301&filename=1022749221.nh

[14] S.J. Zhao, S.L. Ma, M.C. Wang, Super Parameter Optimization of ELM by Artificial Ecosystem-Based Optimization with Crowding Forward-Backward and Backtracking Tips. Control and Decision, (2022) https://kns.cn-ki.net/kcms/detail/21.1124.TP.20220301.0948.005.html

[15] D. Oh, J. Race, S.O.A.B. Koo, Burst Pressure Prediction of API 5L X-Grade Dented Pipelines Using Deep Neural Network, Journal of Marine Science and Engineering, (2020) 766.

[16] R.C.C. Silva, J.N.C. Guerreiro, A.F.D. Loula, A study of pipe interacting corrosion defects using the FEM and neural networks, ADVANCES IN ENGINEERING SOFTWARE, (2007) 868-875.

[17] V. Chauhan, Advances in interaction rules for corrosion defects in pipelines. In: Proceedings of the International Gas Research Conference, (2004). Vancouver, Canada.

[18] Y. Xü, A. Fenerci, O.A. Øiseth, T. Moan, Efficient prediction of wind and wave induced long-term extreme load effects of floating suspension bridges using artificial neural networks and support vector machines(Article), OCEAN ENGINEERING, (2020) 107888.

[19] A.J. Smola, B. Scholkopf, A tutorial on support vector regression, STATISTICS AND COMPUTING, (2004) 199-222.

[20] A.M. Andrew, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, KYBERNETES, (2001) 103-115.

[21] C.J. Evans, T.F. Miller, Failure prediction of pressure vessels using finite element analysis(Article), Journal of Pressure Vessel Technology, Transactions of the ASME, (2015) 51206.

[22] M. Abyani, M.R. Bahaari, M. Zarrin, M. Nasseri, Predicting failure pressure of the corroded offshore pipelines using an efficient finite element based algorithm and machine learning techniques., OCEAN ENGINEERING, (2022) 111382.

[23] S. Sain, The Nature of Statistical Learning Theory, TECHNOMETRICS, (1996) 409.

[24] W. Zhao, L. Wang, Z. Zhang, Artificial ecosystem-based optimization: a novel nature-inspired meta-heuristic algorithm, Neural Computing and Applications, (2020) 9383-9425.

[25] H.R. Tizhoosh, Opposition-based learning: A new scheme for machine intelligence, International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), Vienna, Austria, 2005.

[26] G. Louppe, UNDERSTANDING RANDOM FORESTS: FROM THEORY TO PRACTICE, University of Liège, 2014.

[27] Y. Mahmutoglu, K. Turk, Positioning of leakages in underwater natural gas pipelines for time-varying multipath environment., OCEAN ENGINEERING, (2020).

[28] X.A. Liu, M.A. Xia, D.A. Bolati, J.A. Liu, Q.A. Zheng, H.A.H.C. Zhang, An ANN-based failure pressure prediction method for buried high-strength pipes with stray current corrosion defect., Energy Science & Engineering, (2020) 248-259.

[29] B.M. Wilamowski, H. Yu, Improved Computation for Levenberg-Marquardt Training., IEEE transactions on neural networks, (2010) 930-937.

[30] J. Li, Y. Wu, Improved Sparrow Search Algorithm with the Extreme Learning Machine and Its Application for Prediction, NEURAL PROCESSING LETTERS, (2022) 4189-4209. https://doi.org/10.1007/s11063-022-10804-x

[31] S. Mirjalili, Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems(Article), Neural Computing and Applications, (2016) 1053-1073.