Structural Durability & Health Monitoring

Tech Science Press

https://doi.org/10.32604/sdhm.2026.077489

ARTICLE

Residual Strength Prediction of Corroded Pipelines Based on Sparrow Search Algorithm-Optimized Kernel Extreme Learning Machine

Zixuan Zong[1], Tingting Long[1], Huaqing Dong[1,*], Guoqiang Huang[1], Xiao Meng[1] and Mohammadamin Azimi[2]

[1]Sinopec Sichuan-East Gas Transmission Branch, Wuhan, China
[2]GeoEngineers Inc., San Diego, CA, USA

*Corresponding Author: Huaqing Dong. Email: dhq8888@126.com

Received: 10 December 2025; Accepted: 28 January 2026; Published: 18 May 2026

ABSTRACT: This paper proposes a novel approach for predicting the residual strength of corroded pipelines by combining the Kernel Extreme Learning Machine (KELM) with Sparrow Search Algorithm (SSA) optimization. The proposed SSA-KELM model addresses the limitations of traditional evaluation methods and single machine learning models in residual strength prediction. A dataset comprising 80 samples from burst tests and finite element simulations was used to validate the model. Results demonstrate that the SSA-KELM model achieves superior prediction accuracy with a maximum relative error of 13.54% and minimum relative error of 0.20%. The model's mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) are 0.658%, 0.780%, and 4.38%, respectively, significantly outperforming conventional machine learning models and traditional assessment methods. This research provides a reliable tool for evaluating pipeline integrity and maintenance planning.

KEYWORDS: Pipeline corrosion; residual strength prediction; kernel extreme learning machine; sparrow search algorithm; machine learning

1 Introduction

With the continuous development of the global economy and increasing energy demands, oil and gas pipelines have become critical infrastructure for long-distance energy transportation. However, pipeline failures caused by corrosion have emerged as a significant challenge to operational safety and environmental protection [1,2]. Statistics from various international organizations highlight the severity of this issue. According to the Pipeline and Hazardous Materials Safety Administration (PHMSA), corrosion accounts for approximately 25% of pipeline accidents in the United States [3,4]. The European Gas Pipeline Incident Data Group (EGIG) reports that corrosion ranks as the second leading cause of pipeline failures during 2010–2019. More significantly, data from the Alberta Energy Regulator (AER) indicates that corrosion-related failures constitute up to 50% of pipeline incidents [5].

Accurate prediction of residual strength in corroded pipelines is crucial for ensuring safe pipeline operations and preventing catastrophic failures. As corrosion inevitably occurs during pipeline service and cannot be completely prevented even with various protective measures, the ability to accurately assess the remaining strength of corroded pipelines becomes particularly important for integrity management and maintenance planning. Failure to timely identify severely corroded vulnerable segments could lead to severe consequences [6,7]. Traditional methods for evaluating corroded pipeline residual strength include various

standards and criteria such as ASME B31G, Modified ASME B31G, and DNV-RP-F101 [8]. The ASME B31G standard, developed in 1984 and subsequently modified in 1991, 2009, and 2012, is primarily suitable for lower-grade steel pipes but shows significant limitations when applied to high-strength pipelines [9]. While the DNV-RP-F101 criterion demonstrates better accuracy for medium and high-strength steel pipes, it still exhibits considerable conservatism in practical applications [10]. These traditional evaluation methods often require extensive field testing and can be both time-consuming and resource-intensive.

Recent years have witnessed growing interest in applying machine learning approaches to pipeline residual strength prediction [11,12]. Early attempts using artificial neural networks (ANN) by Xu et al. [13] showed promising results in predicting pipeline failure pressure, achieving prediction accuracy with R^2 values exceeding 0.95. Support Vector Machine (SVM) and Extreme Learning Machine (ELM) have also been explored, with varying degrees of success. For instance, Lo et al. [14] applied SVM to predict corrosion defect failure pressure and reported MAPE values around 8%–12%. Recent studies by Ahmed Soomro et al. [15] demonstrated that Random Forest models could achieve prediction errors below 10% for pipeline burst pressure estimation. Meanwhile, finite element analysis has been widely used in residual strength assessment. Karuppanan et al. [16] utilized finite element analysis for estimating burst pressure of corroded pipelines. Capula Colindres et al. [17] investigated the effects of corrosion defect depth on failure pressure predictions using finite element models, finding that defect depth has the most significant impact on burst pressure reduction. Larin et al. [18] conducted statistical estimation of residual strength based on direct finite element simulations.

However, these existing methods face several significant limitations. Traditional assessment standards, while capable of evaluating pipeline residual strength, often yield overly conservative results with considerable errors. Although such conservative assessments contribute to operational safety, they simultaneously reduce operational efficiency. Finite element methods require detailed modeling for different types of pipelines and defects. This makes the process cumbersome and resource-intensive, which limits their practical engineering applications. For machine learning approaches, the SVM model, while suitable for small sample predictions, sometimes exhibits unstable results and requires lengthy training times. The ELM model, despite its fast training speed, suffers from stability issues due to random weight initialization [19]. The Back Propagation Neural Network (BPNN), despite its widespread use, suffers from slow learning speeds and tends to converge to local optima. This local optima convergence occurs because BPNN uses gradient descent optimization, which can get trapped in suboptimal solutions when the error surface contains multiple local minima. The random initialization of weights further exacerbates this problem, as different starting points may lead to different local optima [20]. Moreover, the determination of hidden layer nodes in BPNN typically relies on empirical methods rather than theoretical guidance, significantly impacting the model's prediction performance. While reinforcement learning methods have shown potential in sequential decision-making tasks, their application to pipeline residual strength prediction remains limited due to the static nature of the prediction problem and the lack of clear reward signals [21].

The Kernel Extreme Learning Machine (KELM) has emerged as a promising approach, combining the advantages of kernel functions with rapid training capabilities. However, the manual adjustment of KELM parameters introduces subjective factors that can affect prediction accuracy. To address this limitation, various optimization algorithms have been explored. The recently proposed Sparrow Search Algorithm (SSA) introduced by Xue and Shen [22] offers unique advantages in parameter optimization due to its strong global search capability and fast convergence characteristics.

Given these challenges and opportunities, this study introduces the KELM model for pipeline residual strength prediction. KELM demonstrates superior capabilities in handling nonlinear problems and rapid training speeds, having shown excellent performance across various domains. Although KELM has been

widely applied in various regression tasks such as wind power prediction, financial forecasting, and medical diagnosis, it has not been previously applied specifically to the prediction of corroded pipeline residual strength. Its characteristics make it particularly promising for this application. This research comprehensively addresses current challenges through several innovative aspects:

- Feature extraction: The study employs Kernel Principal Component Analysis (KPCA) to screen corrosion influence factors, effectively achieving dimensionality reduction for nonlinear data.
- KELM optimization for corrosion rate prediction: Particle Swarm Optimization (PSO) algorithm is utilized to enhance the traditional KELM model's prediction accuracy.
- Advanced optimization for residual strength prediction: A novel Sparrow Search Algorithm (SSA) is implemented to optimize KELM parameters, further improving prediction accuracy.

The remainder of this paper is organized as follows: Section 2 presents the methodological framework, detailing the SSA-KELM hybrid model's structure and optimization process using an 80-sample dataset. Section 3 provides comprehensive comparative analyses against both machine learning models (SVM, ELM, KELM) and traditional assessment standards (ASME B31G, PCORRC), demonstrating the superior prediction accuracy of SSA-KELM. Section 4 summarizes the key findings and contributions to pipeline integrity assessment. This research advances the field by providing a more accurate and practical approach for corroded pipeline residual strength prediction [23].

2 Methodology

2.1 Model Architecture

The proposed model integrates the SSA with the KELM to predict the residual strength of corroded pipelines. SSA is a novel swarm intelligence optimization algorithm proposed by Xue and Shen in 2020, which simulates the foraging and anti-predation behaviors of sparrows.

In the SSA architecture, the sparrow population is divided into three categories: finders with high energy reserves responsible for discovering food-rich areas, joiners who follow finders to forage, and scouts who can detect dangers. Throughout the optimization process, the proportion of these roles remains unchanged, although individual sparrows may switch roles and update their positions in a timely manner based on their energy reserves. The key parameters of SSA, including population size, proportion of discoverers, and safety threshold, are typically set based on empirical guidelines from the original literature. In this study, following the recommendations of Xue and Shen [22], we set the population size to 30, the proportion of discoverers to 20%, and the safety threshold ST to 0.8. The following Eq. (1)–(3) are the expressions for updating the positions of finders, joiners, and scouts, respectively.

$$X_{i,j}^{N+1} = \begin{cases} X_{i,j}^{N} \cdot exp\left(-\dfrac{i}{\beta \cdot iter_{\max}}\right), & \varphi < ST \\ X_{i,j}^{N} + Q \cdot L, \varphi \geq ST \end{cases} \tag{1}$$

In the Eq. (1), N represents the current iteration number of the i-th sparrow; $X_{i,j}$ represents the position of the i-th sparrow individual in the j-th dimension; β is a random number in the interval $[0, 1]$; $iter_{\max}$ is the maximum number of iterations; Q is a random number following normal distribution; L is a $1 \times D$ matrix with all elements being 1; φ is defined as the sparrow's alarm value, with a range of $[0, 1]$; ST is the sparrow's safety threshold value, with a range of $[0.5, 1]$. When $\varphi < ST$, it indicates that the sparrow's environment is safe without predators nearby; when $\varphi \geq ST$, it indicates that some sparrows in the population have detected predators, at which point they will emit alarm signals and lead other sparrows to fly toward safe areas.

$$X_{i,j}^{N+1} = \begin{cases} Q \cdot exp\left(\dfrac{X_{worst}^{N} - X_{i,j}^{N}}{i^2}\right), if\, i > \dfrac{n}{2} \\ X_{s}^{N+1} + \left|X_{i,j}^{N} - X_{s}^{N+1}\right|, otherwise \end{cases} \tag{2}$$

In the Eq. (2), X_{worst}^{N} represents the global worst position at iteration N; n represents the number of sparrows in the population; X_{s}^{N+1} represents the optimal position of the discoverers in the sparrow population after the $N+1$-th iteration; A is a $1 \times D$ matrix with elements randomly assigned as 1 or -1, satisfying $A^{+} = A^{T}\left(AA^{T}\right)^{-1}$; when $i > \frac{n}{2}$, it indicates that the joiner sparrow i has not found food in time and has a poor fitness value, in which case joiner i will choose to forage in other areas; conversely, joiner i will replace the position of a certain discoverer, thereby updating its position.

$$X_{i,j}^{N+1} = \begin{cases} X_{best}^{N} + \lambda |X_{i,j}^{N} - X_{best}^{N}|, if\ f_i > f_m \\ X_{i,j}^{N} + \theta \cdot \left(\dfrac{|X_{i,j}^{N} - X_{worst}^{N}}{(f_i - f_w) + \varphi}\right), if\ f_i = f_m \end{cases} \tag{3}$$

In the Eq. (3), X_{best}^{N} represents the global optimal position at iteration N; λ and θ are step factors, where λ is a random number following normal distribution and $\theta \in [-1, 1]$; φ is a minimal constant to prevent division by zero; f_i, f_m, and f_w represent the current sparrow individual's fitness value, the global optimal fitness value, and the worst fitness value, respectively. When $f_i > f_m$, it indicates that sparrow i is located at the edge of the population and is at risk of predator attack, in which case they will choose to fly to safer areas to obtain better positions; when $f_i = f_m$, it indicates that sparrows in the middle of the population have encountered danger, and they will avoid danger by moving closer to other sparrows.

KELM has been widely applied in regression prediction due to its powerful learning capability and generalization performance. KELM enhances model prediction performance while retaining the advantages of ELM. For predicting the residual strength of corroded pipeline defects, based on the single hidden layer feedforward neural network ELM, the SSA is utilized to optimize the regularization coefficient and kernel function parameters of KELM, thereby further improving KELM model's convergence speed, generalization performance, and prediction accuracy. In light of this, this study employs MATLAB software to conduct predictive research on the residual strength of corroded pipeline defects using the SSA-KELM hybrid model. The prediction flowchart and specific steps for the residual strength of corroded pipeline defects based on the SSA-KELM hybrid model are displayed in Fig. 1 and detailed below.

KELM, with its powerful learning capability and generalization performance, not only mitigates the instability caused by random initialization of weights and thresholds in ELM model training but also maintains the excellent training speed characteristic of ELM while enhancing model stability and nonlinear mapping capability.

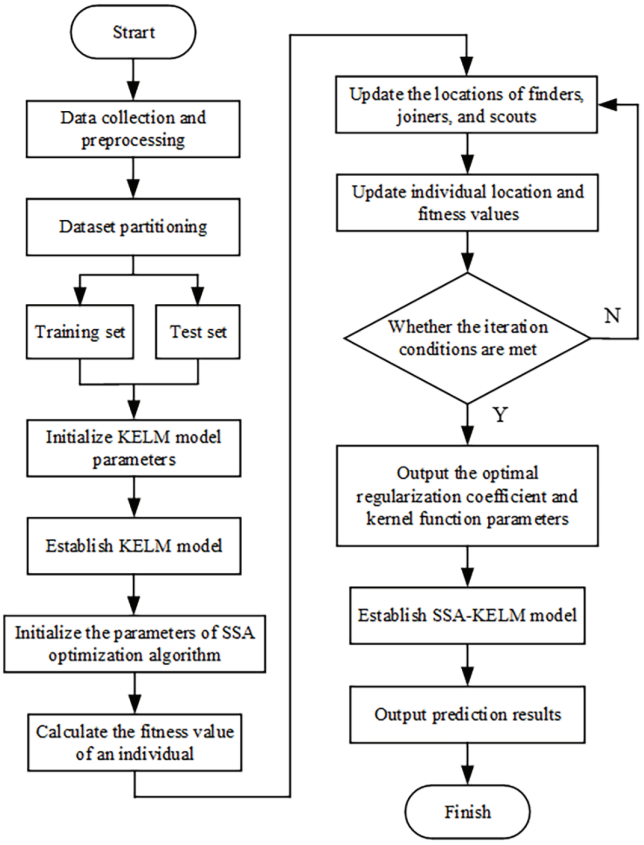

Figure 1: The prediction flowchart for the residual strength of corroded pipeline defects based on the SSA-KELM hybrid model.

KELM incorporates the regularization coefficient C and kernel function Ω_{ELM} into the ELM network. In this study, the Radial Basis Function (RBF) kernel, known for its strong generalization capability, is selected. The kernel function expression K and the output weight β' are given by Eq. (4) and (5):

$$K\left(x_i, x_j\right) = \exp\left(-\frac{||x_i, x_j||^2}{2\sigma^2}\right) \tag{4}$$

$$\beta' = H^T\left(\frac{I}{C} + \Omega_{ELM}\right)^{-1} T \tag{5}$$

where σ represents the width coefficient of the radial basis function, I denotes the identity matrix, C is the regularization parameter, $\Omega_{ELM} = HH^T$, and $\Omega_{i,j} = h\left(x_i\right) h(x_j) = K(x_i, x_j)$.

Based on the above formulas, the output function of the KELM network is expressed as Eq. (6):

$$y\left(x\right) = h\left(x\right)\beta = H\beta = HH^T\left(HH^T + \frac{1}{C}\right)^{-1} T = \begin{bmatrix} K\left(\mathrm{x}, x_1\right) \\ \vdots \\ K\left(\mathrm{x}, x_Q\right) \end{bmatrix}\left(\Omega_{ELM} + \frac{I}{C}\right)^{-1} T \tag{6}$$

For predicting the residual strength of corroded pipeline defects, based on the single hidden layer feedforward neural network ELM, the SSA is utilized to optimize the regularization coefficient C and kernel function parameters γ of KELM, thereby further improving KELM model's convergence speed,

generalization performance, and prediction accuracy. In light of this, this study employs MATLAB software to conduct predictive research on the residual strength of corroded pipeline defects using the SSA-KELM hybrid model. The prediction flowchart and specific steps for the residual strength of corroded pipeline defects based on the SSA-KELM hybrid model are shown in Fig. 1 and detailed below.

Step 1: Collect residual strength data of corroded pipeline defects from literature and conduct correlation analysis on factors affecting the residual strength of oil and gas pipelines. Construct a sample dataset following correlation analysis to achieve dimensionality reduction. Divide the dimensionally reduced dataset into training and testing sets with an 80%:20% ratio (64 samples for training and 16 samples for testing).

Step 2: Initialize the regularization coefficient and kernel function parameters of the KELM model and establish the KELM model structure.

Step 3: Initialize the relevant parameters of the SSA algorithm, including: the number of sparrows in the population n, maximum number of iterations T, proportion of finders in the population, alarm value R_2 and safety threshold value ST, dimensions of variables D.

Step 4: Construct the SSA-KELM regression prediction model, calculate the fitness value of each sparrow individual, and identify the optimal and worst sparrow individuals through comparative ranking.

Step 5: Update the positions of finders, joiners, and scouts in the sparrow population using Eqs. (1)–(3).

Step 6: Terminate the SSA optimization process when the iteration count reaches the maximum iteration number and output the optimal regularization coefficient and kernel function parameters. Otherwise, repeat steps 3 through 5.

Step 7: Establish the SSA-KELM model using the optimized parameters and execute the network to obtain prediction results.

2.2 Data Processing and Model Development

The dataset employed in this investigation comprises 80 samples gathered from pipeline failure experiments and finite element simulations, consisting of 67 specimens from burst tests and 13 from numerical simulations. The combination of experimental burst test data and finite element simulation results is a common practice in pipeline integrity research [24]. Burst tests provide highly reliable ground-truth data but are expensive and limited in quantity. Finite element simulations, when properly validated against experimental results, can supplement the dataset by covering parameter ranges not available in physical tests. In our dataset, the finite element simulation cases were validated against published experimental results, showing prediction errors within 5% for burst pressure. The potential modeling error from finite element simulations is mitigated by using only well-validated cases and by the learning algorithm's ability to generalize from the combined dataset. Each sample incorporates nine parameters: steel grade, diameter, wall thickness, tensile strength, yield strength, defect length, defect width, defect depth, and failure pressure. The factors influencing the residual strength (failure pressure) of corroded pipelines can be categorized into two primary aspects: pipeline intrinsic parameters and defect characteristics.

Prior to predicting the residual strength of corroded pipelines, it is essential to examine the contributing factors. In data-driven models, existing datasets may contain numerous parameters, yet not all variables contribute positively to model training. In this study, we employ a two-stage feature analysis approach. First, Spearman correlation analysis is used to identify and remove features with very low correlation to the target variable (failure pressure). This statistical method is chosen because it can capture monotonic relationships without assuming linearity, making it suitable for identifying irrelevant features. Subsequently, the selected features are used directly as inputs to the model. While Kernel Principal Component Analysis (KPCA) can be used for nonlinear dimensionality reduction in cases with high-dimensional correlated features, in our

case with only five selected input features, direct feature selection based on Spearman correlation proved sufficient. The Spearman coefficient is calculated as follows:

$$yr_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)} \tag{7}$$

where r_s represents the Spearman correlation coefficient between two variables, d_i denotes the difference between the ranks of paired values, n indicates the total number of samples for each variable. The correlation interpretation is as follows: When $r_s > 0$, it indicates a positive correlation between variable pairs; When $r_s < 0$, it signifies a negative correlation between variable pairs; When $r_s = 0$, it suggests no correlation exists between variable pairs.

The precise interpretation of correlation coefficient r values is presented in Table 1.

Table 1: Comparison table of correlation coefficient and correlation degree.

Correlation Coefficient	Correlation Degree
$\lvert r_s\rvert = 0$	Completely unrelated
$0 < \lvert r_s\rvert \leq 0.2$	Very weak correlation
$0.2 < \lvert r_s\rvert \leq 0.4$	Weak correlation
$0.4 < \lvert r_s\rvert \leq 0.6$	Moderate correlation
$0.6 < \lvert r_s\rvert \leq 0.8$	Strong correlation
$0.8 < \lvert r_s\rvert \leq 1$	Very strongly correlated
$\lvert r_s\rvert = 1$	Completely related

Analyze all influencing factors, and the analysis results are revealed in Fig. 2.

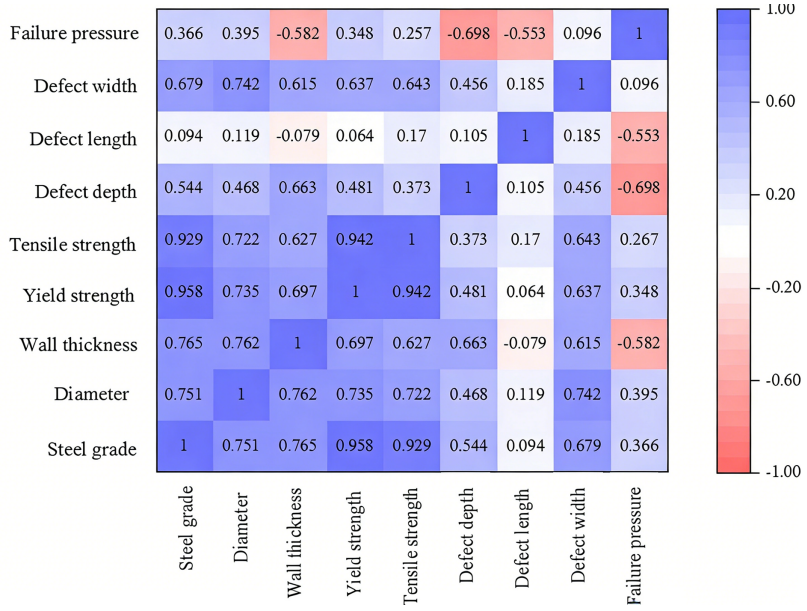

Figure 2: Spearman heatmap of factors affecting pipeline residual strength.

Regarding the correlations among various influencing factors, pipeline steel grade exhibits exceptionally strong correlations with yield strength ($|r_s| = 0.958$) and tensile strength ($|r_s| = 0.929$), thus enabling steel grade to substitute for both yield and tensile strength parameters. In examining the relationships between individual factors and pipeline residual strength, defect width demonstrates negligible correlation ($|r_s| = 0.096$), permitting its exclusion from consideration in this investigation. The threshold value of 0.3 was established based on commonly accepted guidelines in correlation analysis, where $|r_s| < 0.3$ indicates weak or negligible correlation. This threshold ensures that only features with at least weak-to-moderate correlation with the target variable are retained, while removing features that contribute minimal predictive information. Similar thresholds have been adopted in previous pipeline integrity studies [25]. Accordingly, the machine learning model incorporates pipeline steel grade, diameter, wall thickness, defect depth, and defect length as input parameters, while failure pressure serves as the output variable.

Given the disparate dimensions and orders of magnitude in the input-output data for predicting corroded pipeline residual strength, normalization processing was implemented to circumvent numerical issues that could impede model training efficiency. The normalization function maps all data to values within the [−1, 1] interval, as expressed in Eq. (8). It should be noted that the normalized values shown in Table 2 fall within the range [−1, 0] because the original data values in these samples happen to be closer to the minimum values of their respective ranges, resulting in normalized values in the lower portion of the [−1, 1] interval. This is a natural consequence of the data distribution rather than an error in the formula. The normalization function is expressed in Eq. (8), with partial results presented in Table 2.

$$X' = 2 \cdot \frac{X - X_{\min}}{X_{\max} - X_{\min}} - 1 \tag{8}$$

where X' denotes the normalized data; X represents the original data before normalization; $X_{\min}$ signifies the minimum value in the original dataset; $X_{\max}$ indicates the maximum value in the original dataset.

Table 2: Normalization results of pipeline residual strength data.

Number	Steel Grade	Diameter	Wall Thickness	Defect Depth	Defect Length	Failure Pressure
1	−1.0000	−0.9967	−0.9684	−0.9178	−0.8188	−0.3230
2	−1.0000	−0.9969	−1.0000	−0.8270	−0.9493	−0.3184
3	−1.0000	−0.9975	−0.9477	−0.9106	−0.8659	−0.2564
4	−1.0000	−0.9988	−0.9744	−0.9870	−1.0000	−0.2235
5	−1.0000	−0.9993	−0.9537	−0.7462	−0.7826	−0.2199
6	−1.0000	−1.0000	−0.9878	−0.8443	−0.9746	−0.2162
7	−1.0000	−0.9976	−0.9497	−0.8298	−0.9891	−0.1241
8	−1.0000	−0.9981	−0.9562	−0.9913	−0.9783	−0.1068
9	−1.0000	−0.9997	−0.9623	−0.9077	−0.8986	−0.0894
10	−1.0000	−0.4587	−0.4607	−0.5443	−0.5442	−0.6688
11	−1.0000	−0.4587	−0.4607	−0.4578	−0.8295	−0.1788
12	−1.0000	−0.4587	−0.4607	−0.5443	−0.8152	−0.0420
13	−0.7895	−0.8924	−0.9379	−0.6943	−0.9022	−0.5858
14	−0.7895	−0.8940	−0.5180	−0.9077	−0.9710	0.4745
15	−0.7895	−0.8930	−0.5046	−0.8371	−0.8623	0.5100

3 Results and Discussion

3.1 Evaluation Metrics

To assess the practical application effectiveness of SVM, ELM, KELM, PSO-KELM, and SSA-KELM models, this research employs Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination ($R2$) as evaluation metrics. Lower values of MAE, RMSE, and MAPE indicate better performance, with zero representing perfect prediction; $R2$ maximizes at 1, whereby values approaching 1 suggest superior fitting accuracy. These evaluation metrics are expressed in Eqs. (9)–(12).

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\widehat{y_i} - y_i| \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\widehat{y_i} - y_i)^2} \tag{10}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\widehat{y_i} - y_i}{y_i}\right| \times 100\% \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(\widehat{y_i} - y_j\right)^2}{\sum_{i=1}^{N}\left(\widehat{y_i} - \overline{y}_j\right)^2} \tag{12}$$

3.2 Parameter Settings for the Model

Based on the data processing and model development, the model ultimately incorporates steel grade, diameter, wall thickness, defect depth, and defect length as input parameters, while residual strength functions as the output variable. The dataset was randomly partitioned at an 80%:20% ratio, yielding 64 samples for training (80%) and 16 for testing (20%). No separate validation set was used due to the limited sample size; instead, the optimization algorithms (PSO and SSA) use the training error as the fitness function.

Regarding single models (SVM, ELM, and KELM), SVM and KELM share similarities in requiring adjustment of penalty coefficient C and kernel function coefficient γ during training. The hyperparameters $C = 100$ and $\gamma = 5$ for the baseline SVM and KELM models were determined through grid search over the ranges $C \in [0.1, 1000]$ and $\gamma \in [0.01, 100]$, with 5-fold cross-validation on the training set. These values provided the best cross-validation performance among the tested combinations. It should be noted that different hyperparameter settings can significantly affect prediction results; thus, the SSA-optimized KELM model was developed to automatically determine optimal parameters rather than relying on manual tuning. The ELM model necessitates only hidden layer node configuration, which was optimized to 20 nodes through analysis.

To ensure prediction accuracy, this research implements: Population size: 30; Maximum iterations: 100; KELM regularization parameter C range: [0.01, 500]; Kernel function parameter γ range: [0.01, 500].

3.3 Comparative Analysis of Results

The comparative analysis between SSA-KELM predictions and both traditional single models (SVM, ELM, KELM) and conventional standards validates the robustness and accuracy of SSA-KELM in predicting corroded pipeline residual strength.

3.3.1 Comparative Analysis of Prediction Results between SSA-KELM and Single Models (SVM, ELM, and KELM)

The data was applied to train and evaluate SVM, ELM, KELM, and SSA-KELM models. Table 3 and Fig. 3 present the prediction outcomes, while Fig. 4 illustrates relative errors. Table 4 displays prediction accuracy metrics, and Fig. 5 demonstrates the fitting degrees.

Table 3: Prediction results of different models.

Test Sample	Actual Value/MPa	SVM		ELM		KELM		SSA-KELM	
		Predictive Value/MPa	Relative Error/%	Predictive Value/MPa	Relative Error/%	Predictive Value/MPa	Relative Error/%	Predictive Value/MPa	Relative Error/%
1	16.10	16.51	3.66	19.13	18.82	16.88	4.84	16.92	5.09
2	18.06	15.82	12.40	17.36	3.88	16.97	6.04	17.56	2.77
3	8.05	9.04	12.30	9.41	16.89	9.68	20.25	9.14	13.54
4	24.82	22.67	8.66	25.40	2.34	23.58	5.00	23.71	4.47
5	11.51	11.72	1.82	12.78	11.03	12.72	10.51	11.80	2.52
6	14.20	13.11	7.68	13.67	3.73	13.07	7.96	13.38	5.85
7	23.42	21.56	7.94	20.81	11.14	21.08	9.99	23.01	1.75
8	24.91	23.87	4.18	25.24	1.32	24.58	1.32	24.96	0.20
9	21.75	22.88	5.20	22.65	4.14	20.71	4.78	22.43	3.13
10	5.20	9.34	79.62	7.78	49.62	6.20	19.23	5.78	11.15
11	21.21	20.49	3.40	19.18	9.57	17.62	16.93	19.40	8.53
12	17.70	15.23	13.95	15.17	14.29	16.67	5.82	17.23	2.66
13	24.37	23.73	2.63	23.97	1.64	22.01	9.68	25.07	2.87
14	22.00	21.21	3.59	21.24	3.45	19.15	12.95	21.90	0.45
15	15.80	14.74	6.71	12.82	18.86	15.06	4.68	15.46	2.15
16	25.26	24.94	1.27	24.52	2.93	25.89	2.49	24.51	2.97

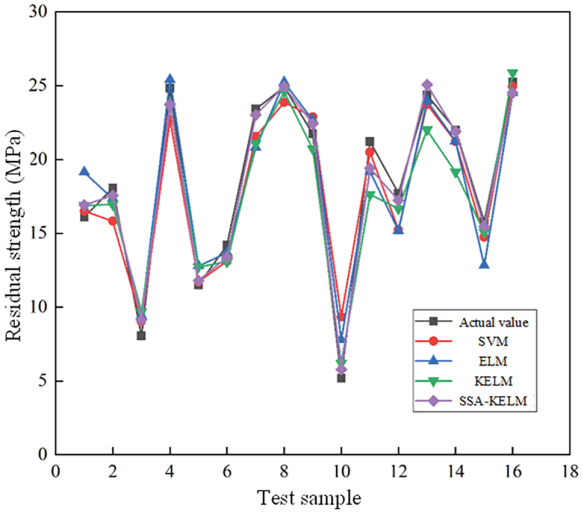

Figure 3: Comparison chart of prediction results.

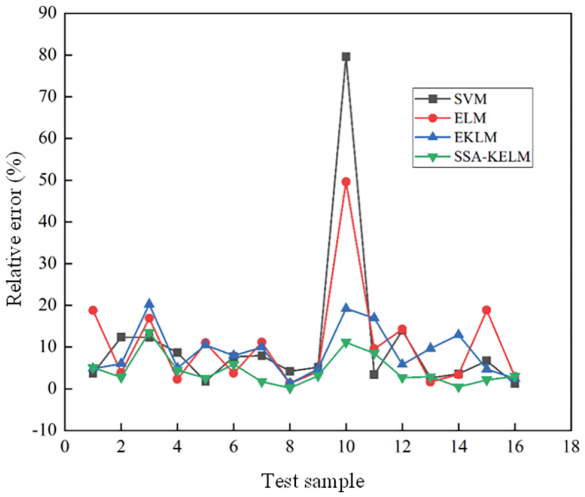

Figure 4: Comparison chart of relative error of prediction results.

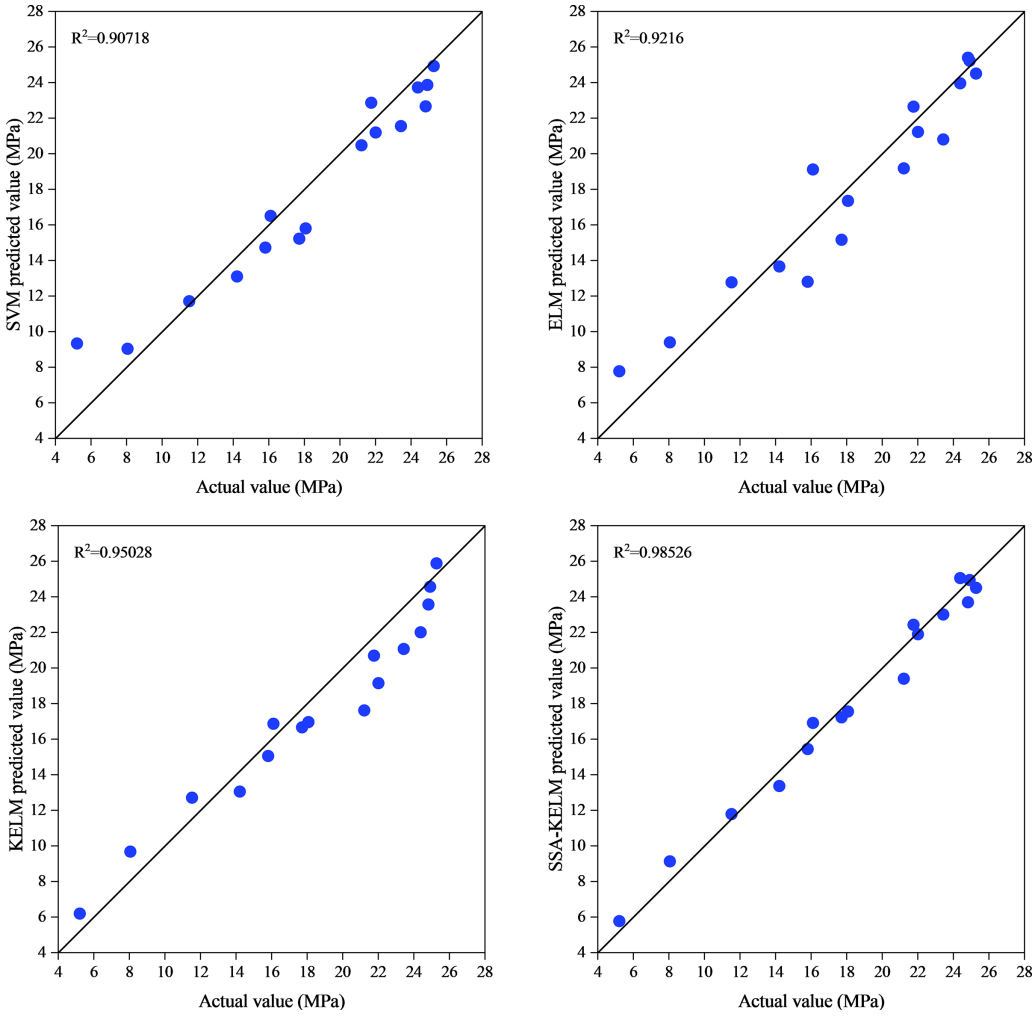

Figure 5: Fitting diagram of the prediction results of the model.

Table 4: Prediction accuracy index of the model.

Model	MAE	RMSE	MAPE	R^2
SVM	1.329	1.656	10.94%	0.90718
ELM	1.454	1.747	10.85%	0.92160
KELM	1.437	1.677	8.90%	0.95028
SSA-KELM	0.658	0.780	4.38%	0.98526

Evidence from Tables 3 and 4, and Figs. 3–5 indicates that the SSA-KELM model surpasses the other three single models in both stability and prediction accuracy. This superiority stems from single models' reliance on manual parameter adjustment, which introduces substantial subjective factors, thereby limiting their effectiveness in analyzing and resolving complex problems.

Initial comparative analysis of relative errors across the four models reveals the following minimum relative errors: SVM: 1.27%; ELM: 1.32%; KELM: 1.32%; SSA-KELM: 0.2%. And maximum relative errors: SVM: 79.62%; ELM: 49.62%; KELM: 20.25%; SSA-KELM: 13.54%. The exceptionally high relative error of 79.62% observed in the SVM model for test sample 10 (actual value 5.20 MPa, predicted 9.34 MPa) warrants discussion. This outlier can be attributed to several factors: (1) the actual failure pressure of 5.20 MPa is the lowest in the test set and lies at the boundary of the training data distribution, making extrapolation challenging; (2) SVM with fixed hyperparameters may not adequately capture the nonlinear relationship in this extreme region; (3) the sample may have unique defect characteristics not well-represented in the training set. The SSA-KELM model significantly reduced this error to 11.15% through automatic parameter optimization, demonstrating its superior generalization capability. Although SVM exhibits marginally lower minimum relative error compared to ELM and KELM, it exceeds SSA-KELM by 1.07%. The difference between maximum and minimum relative errors is 78.35% for SVM and 48.3% for ELM. This indicates that the SVM and ELM models are not stable in predicting the remaining strength of corroded pipelines. For KELM and SSA-KELM models, the difference between the maximum relative error and the minimum relative error is 18.93% and 13.34%, respectively, which is lower than SVM and ELM models.

Table 4 demonstrates that SSA-KELM achieves lower MAE, RMSE, and MAPE values compared to single models, with MAE and RMSE reductions of 0.671 and 0.876 respectively vs. SVM, and MAPE reduction of 4.52% vs. KELM. Fig. 5 reveals R^2 values of 0.90718, 0.92160, 0.95028, and 0.98526 for the respective models, establishing a stability ranking of SVM < ELM < KELM < SSA-KELM. In conclusion, for predicting corroded pipeline residual strength, the SSA-KELM hybrid model consistently outperforms the three single models in both stability and prediction accuracy.

3.3.2 Comparative Analysis of SSA-KELM Model and Evaluation Criteria

Historically, evaluation standards were used to calculate pipeline residual strength; however, these standards exhibit inherent conservatism and are not applicable to all pipeline types. To address the limitations of evaluation standards, this study compares the pipeline residual strength calculated by the SSA-KELM model and evaluation standards with actual values to further investigate the accuracy and applicability of the SSA-KELM model in pipeline residual strength prediction.

ASME B31G is suitable for lower-grade pipelines, while PCORRC is applicable to higher-grade pipelines. Therefore, ASME B31G criteria and PCORRC method were selected for comparative study with the SSA-KELM model, using 16 sample datasets previously predicted by the SSA-KELM model. To enhance the persuasiveness of research results, relative errors of different methods were calculated, where smaller relative

errors indicate higher computational accuracy. Table 5 presents the calculation results of the SSA-KELM model and evaluation standards, Fig. 6 shows the comparison of prediction results between the SSA-KELM model and evaluation standards, and Fig. 7 illustrates their relative error comparison.

Table 5: SSA-KELM and evaluation criteria calculation results.

Test Sample	Failure Pressure/MPa	Evaluation Criteria Calculation Results/MPa		Model Calculation Results/MPa
		ASME B31G	PCORRC	SSA-KELM
1	16.10	13.06	15.51	16.92
2	18.06	14.44	15.28	17.56
3	8.05	6.68	6.00	9.14
4	24.82	24.09	27.92	23.71
5	11.51	8.62	9.54	11.80
6	14.20	8.17	9.67	13.38
7	23.42	18.83	20.37	23.01
8	24.91	23.90	27.80	24.96
9	21.75	19.10	21.31	22.43
10	5.20	5.80	5.13	5.78
11	21.21	15.47	17.93	19.40
12	17.70	15.38	18.09	17.23
13	24.37	21.15	23.37	25.07
14	22.00	19.22	20.62	21.90
15	15.80	14.34	14.47	15.46
16	25.26	22.89	27.20	24.51

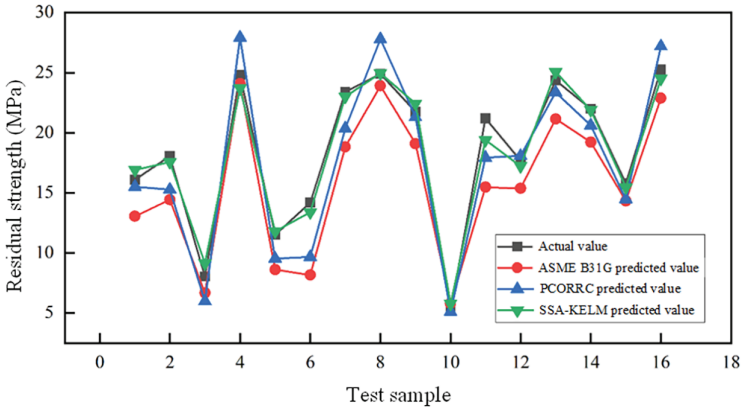

Figure 6: Comparison of prediction results between SSA-KELM and evaluation criteria.

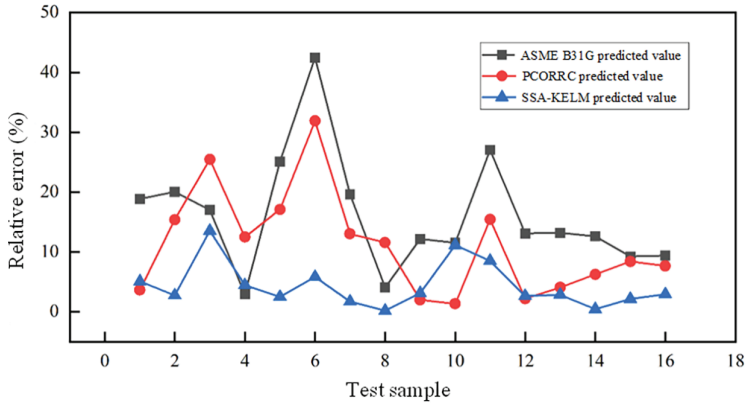

Figure 7: Comparison chart of relative error between SSA-KELM and evaluation criteria.

Analysis of Table 5, Figs. 6 and 7 reveal that the SSA-KELM model's prediction results demonstrate significantly higher accuracy than the evaluation standards, and the prediction results of evaluation standards are almost all lower than actual values, verifying the conservative nature of evaluation standards in calculating pipeline residual strength. Calculation of their prediction results' relative errors shows that the maximum relative errors of ASME B31G method, PCORRC method, and SSA-KELM model are 42.46%, 31.90%, and 13.54%, respectively, while their minimum relative errors are 2.94%, 1.35%, and 0.20%, respectively. The differences between maximum and minimum relative errors for ASME B31G and PCORRC are 39.52% and 30.55%, respectively, sufficiently demonstrating the instability of residual strength results calculated through evaluation standards. In conclusion, the SSA-KELM model yields the best prediction results, followed by the PCORRC method, while the ASME B31G method produces the largest errors. This further validates the high accuracy and applicability of the SSA-KELM model in predicting corroded pipeline residual strength.

3.4 Practical Engineering Considerations

The feasibility of deploying the SSA-KELM model in real engineering scenarios is an important consideration. In terms of computational efficiency, the SSA-KELM model requires approximately 15–30 s for training on the 64-sample dataset using a standard personal computer (Intel i7 processor, 16 GB RAM). Once trained, the prediction for new samples is nearly instantaneous (<0.1 s per sample), making it suitable for real-time applications in pipeline monitoring systems.

Regarding the applicability to different corrosion types, the current model was developed and validated primarily using data from general corrosion defects. The dataset includes both uniform corrosion and localized corrosion cases, though the distinction between pitting and uniform corrosion was not explicitly modeled as a separate feature. For pipelines with predominantly pitting corrosion or other specialized corrosion patterns (such as stress corrosion cracking or microbiologically influenced corrosion), additional calibration with relevant datasets may be necessary. Future work could extend the model by incorporating corrosion type as a categorical input feature, provided sufficient training data for each corrosion type is available [26].

4 Conclusions

This research employs a novel swarm intelligence optimization algorithm (SSA) to optimize the KELM model for accurate prediction of corroded pipeline residual strength. The main conclusions are as follows:

(1) The SSA-KELM prediction model achieves MAE, RMSE, and MAPE values of 0.658%, 0.780%, and 4.38%, respectively. Compared to the SVM model (MAE = 1.329, RMSE = 1.656), this represents reductions of 50.5% in MAE and 52.9% in RMSE. Compared to the ELM model (MAE = 1.454, RMSE = 1.747), the reductions are 54.7% in MAE and 55.4% in RMSE. Compared to the KELM model (MAE = 1.437, RMSE = 1.677), the reductions are 54.2% in MAE and 53.5% in RMSE. This indicates that the SSA-KELM model demonstrates high prediction accuracy for corroded pipeline residual strength.
(2) The SSA-KELM prediction model achieves an R^2 value of 0.98526, showing improvements of 8.6%, 6.9%, and 3.7% compared to SVM, ELM, and KELM models, respectively. This indicates that the SSA-KELM prediction model exhibits higher stability and robustness than SVM, ELM, and KELM models.
(3) Comparative analysis between the SSA-KELM model and two commonly used evaluation standards, ASME B31G and PCORRC, shows that the SSA-KELM model's maximum and minimum relative errors are 13.54% and 0.20%, respectively, representing reductions of 28.92% and 2.74% compared to the ASME B31G method, and reductions of 18.36% and 1.15% compared to the PCORRC method. This further validates the high accuracy and applicability of the SSA-KELM model in predicting corroded pipeline residual strength.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Study conception and design: Zixuan Zong, Guoqiang Huang, Xiao Meng, Tingting Long; data collection and curation: Zixuan Zong, Tingting Long, Mohammadamin Azimi; analysis and interpretation of results: Zixuan Zong, Tingting Long, Mohammadamin Azimi; draft manuscript preparation: Zixuan Zong, Tingting Long, Huaqing Dong, Guoqiang Huang, Xiao Meng, Mohammadamin Azimi. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Data available on request from the authors. The data that support the findings of this study are available from the corresponding author, HD, upon reasonable request.

Ethics Approval: Not applicable.

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

References

1. Wu M, Xie F, Chen X, Wang D, Sun DX. Research progress and thinking on corrosion failure of buried oil and gas pipelines. Oil Gas Storage Transp. 2022;41(6):712–22. (In Chinese). doi:10.1016/j.engfailanal.2022.106951.
2. Farh HMH, El Amine Ben Seghier M, Zayed T. A comprehensive review of corrosion protection and control techniques for metallic pipelines. Eng Fail Anal. 2023;143(3):106885. doi:10.1016/j.engfailanal.2022.106885.
3. Lam C, Zhou W. Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. Int J Press Vessels Pip. 2016;145:29–40. doi:10.1016/j.ijpvp.2016.06.003.
4. Vitali M, Zuliani C, Corvaro F, Marchetti B, Tallone F. Statistical analysis of incidents on onshore CO_2 pipelines based on PHMSA database. J Loss Prev Process Ind. 2022;77:104799. doi:10.1016/j.jlp.2022.104799.
5. Mahmoodian M, Li CQ. Failure assessment and safe life prediction of corroded oil and gas pipelines. J Petrol Sci Eng. 2017;151(3–4):434–8. doi:10.1016/j.petrol.2016.12.029.
6. Vishnuvardhan S, Murthy AR, Choudhary A. A review on pipeline failures, defects in pipelines and their assessment and fatigue life prediction methods. Int J Press Vessels Pip. 2023;201(1):104853. doi:10.1016/j.ijpvp.2022.104853.
7. Al-Sabaeei AM, Alhussian H, Abdulkadir SJ, Jagadeesh A. Prediction of oil and gas pipeline failures through machine learning approaches: a systematic review. Energy Rep. 2023;10(11):1313–38. doi:10.1016/j.egyr.2023.08.009.

8. Li S, Yang Y, Huang B, Jia Y. Residual strength hybrid prediction of hydrogen-blended natural gas pipelines based on FEM-FC-BP model. Energy. 2025;321(82):135463. doi:10.1016/j.energy.2025.135463.
9. Lu H, Peng H, Xu ZD, Qin G, Azimi M, Matthews JC, et al. Theory and machine learning modeling for burst pressure estimation of pipeline with multipoint corrosion. J Pipeline Syst Eng Pract. 2023;14(3):04023022. doi:10.1061/jpsea2.pseng-1481.
10. Netto TA, Ferraz US, Estefen SF. The effect of corrosion defects on the burst pressure of pipelines. J Constr Steel Res. 2005;61(8):1185–204. doi:10.1016/j.jcsr.2005.02.010.
11. Wang Q, Lu H, Li F, Cheng YF. Advancing LightGBM with data augmentation for predicting the residual strength of corroded pipelines. npj Mater Degrad. 2025;9(1):128. doi:10.1038/s41529-025-00673-9.
12. Wang Q, Lu H. Machine learning methods for predicting residual strength in corroded oil and gas steel pipes. npj Mater Degrad. 2025;9(1):30. doi:10.1038/s41529-025-00573-y.
13. Xu WZ, Li CB, Choung J, Lee JM. Corroded pipeline failure analysis using artificial neural network scheme. Adv Eng Softw. 2017;112(11–12):255–66. doi:10.1016/j.advengsoft.2017.05.006.
14. Lo M, Karuppanan S, Ovinis M. Failure pressure prediction of a corroded pipeline with longitudinally interacting corrosion defects subjected to combined loadings using FEM and ANN. J Mar Sci Eng. 2021;9(3):281. doi:10.3390/jmse9030281.
15. Ahmed Soomro A, Akmar Mokhtar A, B. Hussin H, Lashari N, Lekan Oladosu T, Muslim Jameel S, et al. Analysis of machine learning models and data sources to forecast burst pressure of petroleum corroded pipelines: a comprehensive review. Eng Fail Anal. 2024;155(10):107747. doi:10.1016/j.engfailanal.2023.107747.
16. Karuppanan S, Wahab AA, Patil S, Zahari MA. Estimation of burst pressure of corroded pipeline using finite element analysis (FEA). Adv Mater Res. 2014;879:191–8. doi:10.4028/www.scientific.net/amr.879.191.
17. Capula Colindres S, Méndez GT, Velázquez JC, Cabrera-Sierra R, Angeles-Herrera D. Effects of depth in external and internal corrosion defects on failure pressure predictions of oil and gas pipelines using finite element models. Adv Struct Eng. 2020;23(14):3128–39. doi:10.1177/1369433220924790.
18. Larin O, Potopalska K, Mygushchenko R. Statistical estimation of residual strength and reliability of corroded pipeline elbow part based on a direct fe-simulations. J Serb Soc Comput Mech. 2018;12(1):80–95. doi:10.24874/jsscm.2018.12.01.06.
19. Huang GB, Zhou H, Ding X, Zhang R. Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern Part B Cybern. 2012;42(2):513–29. doi:10.1109/TSMCB.2011.2168604.
20. Wu K, Kim K, Garnett R, Gardner JR. The behavior and convergence of local Bayesian optimization. arXiv:2305.15572. 2023.
21. Lu H, Frank Cheng Y. Artificial intelligence in energy pipelines: opportunities and risks. Engineering. 2025;2(11):1–12. doi:10.1016/j.eng.2025.08.032.
22. Xue J, Shen B. A novel swarm intelligence optimization approach: sparrow search algorithm. Syst Sci Control Eng. 2020;8(1):22–34. doi:10.1080/21642583.2019.1708830.
23. Yang X, Zhou Y, Liu Z, Lu L, Yu T, Liu Y. Residual strength prediction technology of corroded pipeline based on RS-PSO-SVM algorithm. Pet Eng Constr. 2020;46(3):8–12. doi:10.1088/1742-6596/1944/1/012019.
24. Wang Q, Lu H, Li F, Cheng YF. Intelligent prediction of residual strength in blended hydrogen–natural gas pipelines with crack-in-dent defects. npj Mater Degrad. 2025;9(1):161. doi:10.1038/s41529-025-00704-5.
25. Lu H, Xu ZD, Zang X, Xi D, Iseley T, Matthews JC, et al. Leveraging machine learning for pipeline condition assessment. J Pipeline Syst Eng Pract. 2023;14(3):04023024. doi:10.1061/jpsea2.pseng-1464.
26. Sabamehr A, Amani N, Boateng S, Sommer A. Development of the new machine-learning approach in pipeline condition assessment prediction and optimizing rehabilitation strategies. J Facil Manag. 2025;23(5):929–48. doi:10.1108/jfm-06-2024-0077.