Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Feature Sensitivity and Robustness in Corrosion Rate Forecasting: A Comparative Study of Deterministic and Probabilistic Machine Learning Models

Jinlong Hu

hujinlong.1988@163.com

Xi'an University

Xiaoli Zhang

Xi'an University

Shaofei Dong

Xi'an University

Zhifu Yin

Xi'an University

Article

Keywords: Corrosion rate prediction, Extra Trees Regressor (ETR), Improved Bayesian Neural Network (ImprovedBNN), Uncertainty quantification, Feature sensitivity

Posted Date: March 30th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9145430/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Feature Sensitivity and Robustness in Corrosion Rate Forecasting: A Comparative Study of Deterministic and Probabilistic Machine Learning Models

Jinlong Hu[a,*], Xiaoli Zhang[a], Shaofei Dong[a], Zhifu Yin[a]

[a]*Shaanxi Key Laboratory of Surface Engineering and Remanufacturing, Xi'an University, Xi'an 710065, China*

Abstract: Accurate corrosion rate prediction is critical for ensuring industrial infrastructure safety and reducing economic losses, yet existing models often fail to balance precision, uncertainty quantification, and adaptability to dynamic scenarios. This study systematically compared the Extra Trees Regressor (ETR) and an Improved Bayesian Neural Network (ImprovedBNN) across static performance metrics (MSE, MAE, coverage), feature sensitivity to ±10% fluctuations, and statistical robustness via paired t-tests. The results show that ETR outperformed in point prediction (MSE=0.0051) while ImprovedBNN achieved perfect uncertainty coverage (1.000) with enhanced sensitivity to critical features, with both models demonstrating distinct engineering applicability. These findings establish a framework for model selection in corrosion management, reconciling deterministic precision and probabilistic rigor to inform practical decision-making.

Keywords: Corrosion rate prediction, Extra Trees Regressor (ETR), Improved Bayesian Neural Network (ImprovedBNN), Uncertainty quantification, Feature sensitivity

1. Introduction

Corrosion, as an inevitable electrochemical process in industrial systems, poses severe threats to industrial infrastructure integrity, operational safety, and economic sustainability across sectors such as oil and gas, maritime transport, and chemical manufacturing [1-2]. Accurate prediction of corrosion rates is therefore critical for optimizing maintenance schedules, reducing lifecycle costs, and mitigating catastrophic risks [3]. Over the past century, corrosion rate prediction has evolved from empirical observations to data-driven modeling, with each approach offering distinct advantages and limitations.

Early efforts to predict corrosion rates relied heavily on empirical formulas and mechanistic models derived from laboratory experiments. The Tafel equation, proposed by Julius Tafel in 1905, describes the relationship between corrosion current density and electrode potential, forming the basis for electrochemical corrosion rate calculations [4]. This equation, coupled with polarization resistance measurements (via linear polarization resistance, LPR), remains a standard in corrosion testing [5]. However, it assumes uniform corrosion and fails to account for localized phenomena such as pitting or crevice corrosion.

Another foundational approach is empirical models based on environmental factors. For instance, the de Waard-Milliams model quantifies CO_2 corrosion rates in oil and gas pipelines using temperature, pressure, and CO_2 partial pressure, with modifications to include water chemistry and flow effects [6]. Similarly, the ISO 12944 standard provides empirical correlations for atmospheric

* Corresponding author. E-mail: hujinlong.1988@163.com

corrosion rates based on pollutant concentrations (e.g., chloride, sulfur dioxide) and exposure time [7]. While these models are widely adopted in industry, they are limited by their reliance on idealized conditions and inability to capture complex, non-linear interactions between variables [8-9].

Over the past decade, data-driven predictive models have emerged as powerful tools to complement traditional methods. Machine learning (ML) algorithms, including ensemble methods (e.g., Random Forest, Extra Trees) and neural networks, have demonstrated superior performance in capturing complex, non-linear relationships between environmental variables (e.g., temperature, pressure, chemical concentrations) and corrosion behavior [10-11].

Ensemble learning methods have gained prominence for their robustness and interpretability. Random Forest (RF) models, which aggregate predictions from multiple decision trees, have been applied to predict corrosion rates in marine environments, outperforming linear regression by 20–30% [12]. Extra Trees Regressors (ETR), a variant of RF with randomized split points, have shown even higher accuracy in predicting pipeline corrosion under dynamic flow conditions [13]. Similarly, Gradient Boosting Machines (GBM) have been used to model corrosion in reinforced concrete, integrating variables such as concrete pH, chloride content, and curing time [14].

Neural networks (NNs) have excelled in handling high-dimensional, non-linear data. Feedforward neural networks (FNNs) have been employed to predict corrosion rates of low-alloy steels in marine atmospheres, achieving mean absolute errors (MAEs) as low as 0.02 mm/year [15]. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks, which capture temporal dependencies, have been applied to time-series corrosion data from sensors, improving predictions of long-term corrosion trends in offshore platforms [16-17].

Probabilistic frameworks such as Bayesian Neural Networks (BNNs) have gained attention for their ability to quantify prediction uncertainty-a critical feature for risk-informed decision-making in engineering practice [18-19]. BNNs assign probability distributions to model parameters, enabling robust uncertainty estimates that account for data noise and model limitations [20]. For example, BNNs have been used to predict corrosion rates in coastal bridges, providing confidence intervals that guide inspection prioritization [21].

Despite advancements, existing corrosion rate prediction methods face significant challenges. First, adaptability to complex environments remains limited. As Wang et al. [22] pointed out, most ML models perform well under controlled conditions but degrade in real-world settings with dynamic variables (e.g., fluctuating temperature, multi-pollutant interactions). For instance, in the complex coastal environment described by Yan et al. [23], the combined effects of temperature, humidity, sodium chloride particles, and other pollutants form a highly non-linear system that existing models struggle to capture comprehensively.

Second, generalizability to extreme conditions is insufficiently verified. While some studies (e.g., [22]) have optimized model parameters using algorithms like Particle Swarm Optimization (PSO) [24], few have validated performance under extreme operating conditions (e.g., supra-threshold chloride concentrations, synergistic changes in pressure and temperature) [25]. Extreme conditions

are common in industrial scenarios (e.g., acid gas leaks, offshore storms), and their exclusion limits model applicability.

Third, uncertainty quantification is often overlooked or poorly calibrated. Deterministic models (e.g., RF, FNN) provide point predictions but lack explicit uncertainty estimates, hindering risk assessment [26]. While BNNs address this, early implementations suffered from overconfident intervals or high computational costs [27-28]. Recent studies have improved BNN calibration using dropout regularization and mixture loss functions [29], but these advancements remain underexplored in corrosion applications.

Finally, feature interpretability is lacking in "black-box" models. Deep NNs, for example, often fail to identify critical corrosion drivers (e.g., chloride vs. oxygen), limiting their utility for material selection or process optimization [30-31].

These challenges converge into a critical research gap: the lack of a comprehensive framework for evaluating corrosion rate prediction models that integrates static performance metrics, dynamic responsiveness to feature fluctuations, and alignment with practical engineering needs. Addressing this gap is essential to guide model selection, enhance the translatability of research findings, and ultimately improve corrosion management strategies in industrial settings.

Despite significant advancements in corrosion rate prediction models, a critical gap persists in reconciling three interconnected demands: high-precision point predictions, reliable uncertainty quantification, and clear alignment with engineering scenario requirements. Existing literature tends to prioritize one dimension at the expense of others: deterministic models (e.g., ensemble methods) often excel in prediction accuracy but lack rigorous uncertainty communication, while probabilistic frameworks (e.g., traditional Bayesian neural networks) struggle to balance coverage, interval precision, and sensitivity to critical feature fluctuations.

Furthermore, comparative studies rarely validate model robustness across dynamic scenarios-specifically, how predictions and uncertainty estimates adapt to controlled feature perturbations (e.g., ±10% fluctuations in key variables like CIC). This leaves engineers without clear guidance on model selection for context-specific tasks, such as real-time monitoring versus risk-based planning. Most critically, the link between feature sensitivity patterns and model applicability remains underexplored, hindering the translation of theoretical performance metrics into practical corrosion management strategies.

It is this gap-between fragmented model evaluations, unvalidated dynamic robustness, and ambiguous engineering relevance-that necessitates a systematic comparison of advanced predictive models, with a focus on both static performance and dynamic response to feature fluctuations.

To address the identified research gap, this study aims to answer the following focused research questions, which collectively guide the empirical investigation and structure the analysis:
1) Performance comparison under static conditions: How do advanced predictive models (specifically, Extra Trees Regressor (ETR) and improved Bayesian Neural Networks (BNNs)) differ

in terms of point prediction accuracy (measured by MSE, MAE, RMSE) and uncertainty quantification (assessed via coverage, calibration error, and interval width) when applied to corrosion rate forecasting?

2) Sensitivity to feature fluctuations: How do these models respond to controlled ± 10% perturbations in individual input features (e.g., CIC, TP, temperature), particularly in terms of prediction trend consistency and adaptive adjustment of uncertainty estimates? Do they identify the same set of critical sensitive features, and if not, what implications do these differences hold for model interpretability?

3) Robustness in dynamic scenarios: Can the observed performance patterns (accuracy, uncertainty, sensitivity) of ETR and improved BNNs be validated across typical operating ranges, ensuring their reliability for real-world corrosion management tasks?

These questions are designed to systematically evaluate the trade-offs between model performance metrics, their ability to adapt to dynamic feature fluctuations, and their alignment with practical engineering needs-ultimately providing actionable guidance for model selection in corrosion rate forecasting.

The remainder of this paper is structured to systematically address the research questions outlined above, following a logical progression from methodology to conclusions. Section 2 details the materials, datasets, and analytical frameworks employed in this study. It first describes the corrosion rate dataset, including the selection of input features (e.g., CIC, TP, temperature) and data preprocessing steps. Subsequently, the technical specifications of the compared models — specifically, the architecture of the Extra Trees Regressor (ETR) and improved Bayesian Neural Network (ImprovedBNN) are introduced, including hyperparameter settings and training protocols. Finally, it defines the performance metrics used for evaluation, such as MSE, MAE, RMSE (for point prediction accuracy), and coverage, calibration error, interval width (for uncertainty quantification), as well as the methodology for feature sensitivity analysis and statistical significance testing.

Section 3 (“Results”) presents the empirical findings in a structured manner, aligning with the research questions. It begins with a comparative analysis of model performance under static conditions, followed by an exploration of model responses to controlled feature fluctuations. Statistical validation of observed differences and insights into feature sensitivity patterns are also included, providing objective evidence to support subsequent discussions.

Section 4 (“Discussion”) interprets the results in the context of existing literature, emphasizing the core scientific contributions of this study. It elaborates on the practical implications of model performance differences for engineering applications, addresses the identified methodological limitations, and discusses how these constraints inform future research directions.

Section 5 (“Conclusion”) synthesizes the key findings, reiterates their significance for corrosion rate prediction and engineering practice, and offers concise recommendations for model selection in

specific scenarios, thereby concluding the study with actionable insights.

2. Method

The methodological workflow of this study is visualized in Figure 1, which outlines the sequential steps from data acquisition to result interpretation, ensuring transparency and reproducibility of the analysis.

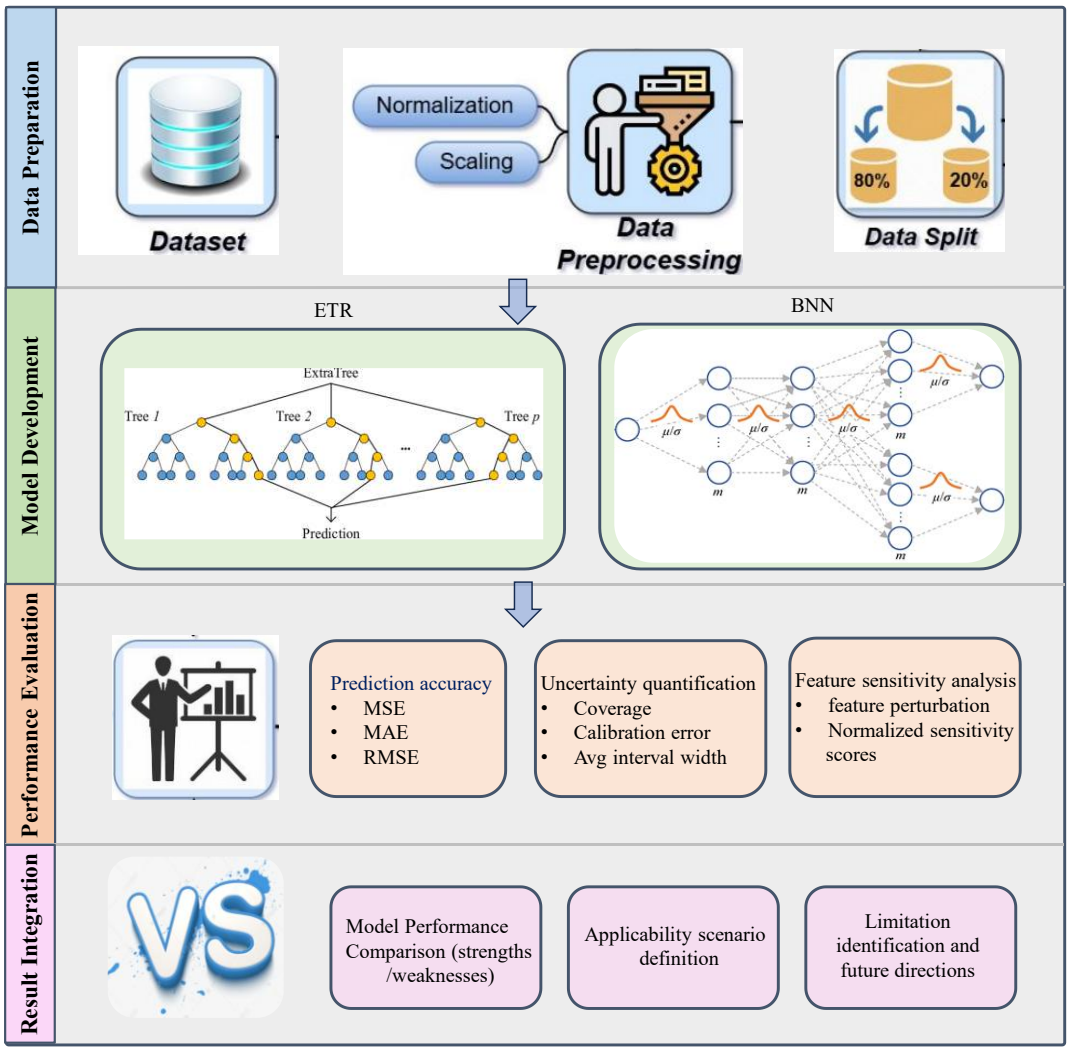


Fig. 1. Schematic overview of the overall framework

This flow diagram illustrates the four core stages of the study: (1) Data Preparation, involving dataset collection, preprocessing (missing value imputation, outlier handling, normalization), and splitting into training/testing subsets; (2) Model Development, including the implementation and optimization of the Extra Trees Regressor (ETR) and Improved Bayesian Neural Network (ImprovedBNN); (3) Performance Evaluation, encompassing assessments of point prediction accuracy (MSE, MAE, RMSE), uncertainty quantification (coverage, calibration error, interval width), feature sensitivity to $\pm 10\%$ perturbations, and statistical validation via paired t-tests; (4) Result Integration (purple), where findings are synthesized to compare model strengths, define applicable scenarios, and identify limitations. Arrows indicate the directional flow of data and analysis.

The framework is designed to systematically address the research questions by linking data characteristics to model behavior, ensuring that each analytical step builds on prior results. Data preparation establishes a consistent foundation for model training, while the evaluation stage employs multi-dimensional metrics to capture both predictive performance and robustness. Statistical tests validate the significance of observed differences, and integration synthesizes these insights into actionable conclusions for corrosion engineering applications.

2.1 Dataset Description and Preprocessing

The experimental data comes from on-site experiments of J55 tubing steel under the environment of oil extraction wells in the Chang qing oilfield air foam test area. The experiments simulated the corrosion rate of J55 tubing steel in Chang qing oilfield in injection wells, production wells under different high temperature and high-pressure conditions. The chemical composition (in wt.%) and mechanical property of the J55 steel, experimental conditions and the amide inhibitor are described in Reference [32].

The corrosion rate data under different conditions were obtained by testing samples of different numbers. Among them, the testing conditions include experimental temperature, total pressure, oxygen content, carbon dioxide pressure, flow rate, whether or not to add corrosion inhibitor and the type of corrosion inhibitor, the mass of the samples before and after the test, and the dimensions of the specimens (length, width and height). The length of the experiment was 168 hours. As a result of these tests, 36 high-quality experimental data were obtained as input data for machine learning modeling.

The study begins with the collection of a comprehensive dataset comprising various influential factors. Each parameter is systematically varied in controlled experiments to simulate real-world conditions in oil wells. The resultant corrosion rates are measured and recorded, providing a robust foundation for training our machine learning models. Due to the high cost, long cycle and complex conditions, there are only 36 experimental data obtained. Data preprocessing was conducted with the method described in our previous work [33].

2.2 Model Specifications and Training Protocols

In this work, two primary models were compared: the Extra Trees Regressor (ETR) and an Improved Bayesian Neural Network (ImprovedBNN), with hyperparameters optimized via 5-fold cross-validation on the training set. Extra Trees Regressor (ETR) was implemented using scikit-learn (v1.2.2), the ETR consisted of 500 decision trees with a maximum depth of 15, minimum samples per leaf set to 5, and a random state of 42 for reproducibility. Splitting criteria were based on mean squared error (MSE), and bootstrap sampling was disabled to reduce variance.

Improved Bayesian Neural Network (ImprovedBNN) was built using PyTorch (v2.0) with probabilistic layers, the network architecture included an input layer (8 neurons), two hidden layers (64 and 32 neurons, respectively) with ReLU activation, and an output layer with a Normal distribution (mean and standard deviation) to model predictive uncertainty. The original BNN's limitations (e.g., overconfident uncertainty) were addressed by: (1) Adding dropout layers (rate=0.2)

after each hidden layer to induce stochasticity; (2) Using a scaled mixture of Gaussians as the loss function to penalize calibration errors; (3) Optimizing via stochastic gradient descent (SGD) with a learning rate of 0.001, momentum of 0.9, and 500 epochs (early stopping applied if validation loss plateaued for 20 epochs).

2.3 Evaluation Metrics and Analytical Methods

Model performance was evaluated across three dimensions: point prediction accuracy, uncertainty quantification, and feature sensitivity, with statistical significance tested to validate observed differences.

2.3.1 Point prediction accuracy

To evaluate the performance of the proposed approach, we employed four standard evaluation metrics for the regression task, mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE, which are described as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y_i\right| \tag{1}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{3}$$

Assuming a set of samples $y = \{y_1, y_2, ..., y_n\}$ and the predicted dataset $y = \{\hat{y}_1, \hat{y}_2, ..., \hat{y}_n\}$, where n is the number of a sample dataset. For training and testing dataset, the smaller the MAE, MAPE and RMSE values, indicating that the model's prediction results are closer to the real experimental results, the better the prediction ability.

2.3.2 Uncertainty Quantification

Uncertainty in corrosion rate predictions arises from multiple sources, including measurement noise in field data (e.g., sensor inaccuracies in chloride ion concentration or temperature readings), incomplete characterization of complex electrochemical reactions, and inherent limitations in model architectures (e.g., over-simplification of nonlinear feature interactions). In engineering practice, unquantified or poorly calibrated uncertainty can lead to overconfident decision-making (e.g., underestimating corrosion risks) or excessive conservatism (e.g., unnecessary maintenance), highlighting the need for rigorous metrics to assess probabilistic prediction reliability. Three complementary metrics were employed to capture different aspects of uncertainty quality, consisting of Coverage, Calibration Error and Avg Interval Width.

2.3.2.1 Coverage

Coverage quantifies the extent to which predicted confidence intervals (CIs) “cover” the true corrosion rates, serving as a fundamental check for the validity of uncertainty estimates. A model

with perfect coverage (value = 1.0) indicates that 95% of true values fall within the 95% CI, reflecting unbiased uncertainty calibration. For a dataset with *n* samples, coverage is calculated as:

$$Coverage = \frac{1}{n}\sum_{i=1}^{n} I\left(y_i \in [\hat{y}_i - 1.96\hat{\sigma}_i, \hat{y}_i + 1.96\hat{\sigma}_i]\right) \quad (4)$$

where I(·) is an indicator function (1 if true value y_i lies within the interval, 0 otherwise), $\hat{y}_i$ is the predicted mean, and $\hat{\sigma}_i$ is the predicted standard deviation.

Coverage is Critical for validating whether uncertainty estimates are neither over- nor under-confident. Low coverage suggests intervals are too narrow (underestimation of uncertainty), while coverage >0.95 may indicate overly wide intervals (wasted precision). In this study, coverage was computed for both ETR and ImprovedBNN to compare their ability to reliably bound true corrosion rates. Statistical tests (paired t-tests) were used to assess whether differences in coverage between models were significant.

2.3.2.2 Calibration Error

Calibration error extends coverage by evaluating the alignment between predicted probabilities and empirical frequencies across all possible error thresholds, not just the 95% CI. A well-calibrated model ensures that, for any threshold τ, the proportion of samples where $|\hat{y}_i - y_i| > \tau$ matches the predicted probability of such errors. For *m* equally spaced thresholds $\tau_1, \tau_2, \ldots, \tau_m$, (ranging from 0 to the maximum observed error), calibration error is:

$$Calibration\ Error = \frac{1}{m}\sum_{k=1}^{m} |\hat{P}(|\hat{y} - y| > \tau_k) - \hat{F}(\tau_k)| \quad (5)$$

Where $\hat{P}(.)$ is the predicted probability of error exceeding τ_k (derived from the model's output distribution), and $\hat{F}(\tau_k)$ is the empirical frequency of such errors in the test set.

Calibration error is essential for scenarios requiring nuanced risk assessment (e.g., prioritizing inspections based on error likelihood). A calibration error of 0 indicates perfect alignment between predictions and observations. In this study, Calibration error was calculated using 50 thresholds (0 to 0.5 mm/year) to compare how well ETR and ImprovedBNN captured the full distribution of prediction errors, particularly in dynamic scenarios with feature fluctuations.

2.3.2.3 Avg Interval Width

While coverage and calibration focus on "reliability", average interval width measures "precision"-the compactness of uncertainty estimates. Narrower intervals (with acceptable coverage) indicate more informative uncertainty quantification, enabling more decisive engineering actions. For 95% CIs, average width is:

$$Avg\ Interval\ Width = \frac{1}{n}\sum_{i=1}^{n} [(\hat{y}_i + 1.96\hat{\sigma}_i) - (\hat{y}_i - 1.96\hat{\sigma}_i)] = \frac{1}{n}\sum_{i=1}^{n} 3.92\hat{\sigma}_i \quad (6)$$

Avg Interval Width is critical for balancing reliability and efficiency. For example, in real-time monitoring, narrow intervals with high coverage allow for precise maintenance scheduling, whereas overly wide intervals reduce actionable insights. In this study, Average Interval Width was tracked across static and dynamic (feature-perturbed) test sets to assess how model precision degraded under feature fluctuations, with ETR and ImprovedBNN compared for both absolute width and width-coverage trade-offs.

Together, these metrics provide a holistic assessment of uncertainty quality. Coverage ensures intervals are reliable, calibration error validates probabilistic consistency, and average width measures practical utility-collectively enabling informed comparisons between deterministic (ETR) and probabilistic (ImprovedBNN) modeling approaches.

2.3.3 Feature Sensitivity Analysis

Feature sensitivity analysis originates from the need to understand how variations in input variables affect model outputs, a critical concern in both machine learning interpretability and engineering decision-making. In corrosion rate prediction, environmental and operational features (e.g., chloride concentration, temperature) often fluctuate in real industrial settings, due to factors like sensor drift, process adjustments, or seasonal changes, and such fluctuations can propagate into prediction errors if models fail to adapt. Traditional "black-box" models often obscure which features drive predictions, limiting trust in their recommendations for corrosion management. Sensitivity analysis addresses this by quantifying the magnitude of output changes in response to controlled input perturbations, revealing both the importance of individual features and the robustness of models to real-world variability.

The sensitivity analysis in this study followed a one-at-a-time (OAT) perturbation protocol, a widely adopted approach for isolating the impact of individual features while holding others constant. This method is particularly suited for corrosion systems, where features often exhibit non-linear interactions but single-variable fluctuations (e.g., $\pm 10\%$ in chloride ion concentration) are common in operational scenarios. For each feature f_k (where $k = 1,2,\ldots,8$) corresponding to CIC, TP, T, etc.), the perturbation range was set to $\pm 10\%$ relative to its mean value (μ_{f_k}) in the test set. This range was chosen based on industrial data showing that typical operational fluctuations in corrosion-related features rarely exceed 10% under normal conditions, ensuring ecological validity. The perturbed values for feature f_k were thus:

$$f_k^{perturbed} = \mu_{f_k} \times (1 \pm 0.10) \tag{7}$$

For each perturbed feature, two key metrics were recorded: $\Delta\hat{y}$ and $\Delta\hat{\sigma}$. $\Delta\hat{y}$ represents the absolute change in corrosion rate predictions, calculated as the difference between predictions with the perturbed feature and predictions with the original feature value:

$$\Delta\hat{y}_k = |\hat{y}(f_k^{perturbed}) - \hat{y}(f_k^{original})| \tag{8}$$

$\Delta\hat{\sigma}$ represents the absolute change in uncertainty estimates, defined similarly for the predicted standard deviation:

$$\Delta\hat{\sigma}_k = |\hat{\sigma}(f_k^{perturbed}) - \hat{\sigma}(f_k^{original})| \quad (9)$$

Furthermore, to conduct cross-feature comparison (accounting for differences in feature scales, e.g., temperature in °C vs. pressure in MPa), a normalized sensitivity score was computed for each feature:

$$S_k = \frac{\Delta\hat{y}_k}{Range(f_k^{perturbed})} = \frac{\Delta\hat{y}_k}{0.20 \times \mu_{f_k}} \quad (10)$$

Here, the denominator represents the total perturbation range (20% of the feature mean), ensuring S_k reflects the "per unit perturbation" impact on predictions. Higher S_k indicates greater sensitivity to that feature.

Feature sensitivity analysis is invaluable in corrosion engineering. First, critical control variables could be identified with feature sensitivity analysis. High-sensitivity features indicate parameters that require strict monitoring to maintain prediction reliability. Second, model robustness can be assessed. Models with stable $\Delta\hat{y}$ and $\Delta\hat{\sigma}$ across perturbations are preferred for dynamic industrial environments. Lastly, it can also guide data collection. Resources can be prioritized for measuring high-sensitivity features with greater precision, reducing overall prediction uncertainty.

In this work, sensitivity analysis was applied to both the ETR and ImprovedBNN. It can be used to compare how each model prioritizes features, e.g., whether both models identify CIC and TP as the most sensitive, validating alignment with corrosion chemistry principles. Then, dynamic robustness will be evaluated with sensitivity analysis. By tracking $\Delta\hat{y}$ and $\Delta\hat{\sigma}$ across all 8 features, we quantified whether ETR's stable predictions or ImprovedBNN's adaptive uncertainty better suited real-world fluctuations. Furthermore, features with high S_k in both models (e.g., temperature) were flagged as critical for operational control, while low-sensitivity features (e.g., pipeline length) indicated areas where measurement costs could be reduced.

2.3.4 Statistical Significance Testing

Statistical significance testing, rooted in frequentist statistics, emerged to address the challenge of distinguishing genuine patterns from random variation in data. In comparative studies of predictive models, raw performance differences (e.g., lower MSE for one model over another) could arise by chance due to sampling variability, especially with finite datasets. Without rigorous statistical validation, such differences might misleadingly inform model selection, undermining the reliability of engineering recommendations. The paired t-test, a cornerstone of this framework, is specifically designed to compare two related samples (e.g., predictions from two models on the same test set), making it ideal for isolating true performance gaps from random noise.

In corrosion rate prediction, model performance metrics (e.g., prediction errors, interval widths) are inherently variable across test samples due to the complexity of corrosion systems and dataset heterogeneity. A paired t-test accounts for this by analyzing the difference between paired observations (e.g., error of Model A vs. error of Model B on the same sample), rather than comparing aggregated metrics directly. This approach controls for sample-specific variability (e.g., a

particularly noisy corrosion measurement), ensuring that observed differences are attributed to model performance rather than data idiosyncrasies.

For paired t-tests, the null hypothesis (H_0) states that the mean difference between paired observations is zero (i.e., no true difference between models), while the alternative hypothesis (H_1) posits a non-zero mean difference (two-tailed test). Given n paired observations (e.g., prediction errors of ETR and ImprovedBNN on n test samples), let $d_i = Error_{Model\ 1,i} - Error_{Model\ 2,i}$ represent the difference for the i-th sample. The test statistic t is calculated as:

$$t = \frac{\overline{d}}{s_d/\sqrt{n}} \tag{11}$$

Where $\overline{d}$ is the mean of the differences ($\overline{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$), s_d is the sample standard deviation of the differences ($s_d = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2}$), n is the number of paired samples.

Paired t-tests are particularly valuable in model benchmarking, metric-specific comparisons and engineering validation. In this work, paired t-tests (two-tailed, α=0.05) were conducted on prediction errors and interval widths between model pairs to validate whether performance differences were non-random, with p-values visualized via heatmaps for clarity.

For each model pair, 7 paired observations were analyzed, and p-values were computed for two-tailed tests. Results were visualized using heatmaps, where cells were color-coded by p-values to intuitively highlight significant vs. non-significant differences.

By integrating statistical significance testing, this study moves beyond descriptive comparisons, providing a rigorous foundation for selecting models that deliver genuine, non-random improvements in corrosion rate prediction.

3. Results

3.1 Comparison of model performance: quantification of accuracy and uncertainty

In the comparison of model basic performance, the precision and uncertainty quantification indicators of different models in the corrosion rate prediction task were analyzed, the result is shown in Figure 2. In terms of precision indicators, for the Mean Squared Error (MSE), the Extra Trees Regressor performed optimally, with an MSE value of 0.0051, which was much lower than that of Gaussian Process Regression (0.0905), Random Forest (0.1059), Original Bayesian NN (0.0420), and Improved Bayesian NN (0.0191). A similar trend was observed for the Mean Absolute Error (MAE). The Extra Trees Regressor had the smallest MAE of 0.0591 among all models. The MAEs of Gaussian Process Regression and Random Forest were relatively high, at 0.2758 and 0.2626 respectively. The MAE of the Original Bayesian NN was 0.1671, while that of the Improved Bayesian NN decreased to 0.0942. The Root Mean Squared Error (RMSE) results also showed that the Extra Trees Regressor led with an RMSE value of 0.0716. The RMSEs of other models were relatively larger. Among them, the RMSE of the Improved Bayesian NN was 0.1381, which was better than that of the Original Bayesian NN (0.2049).

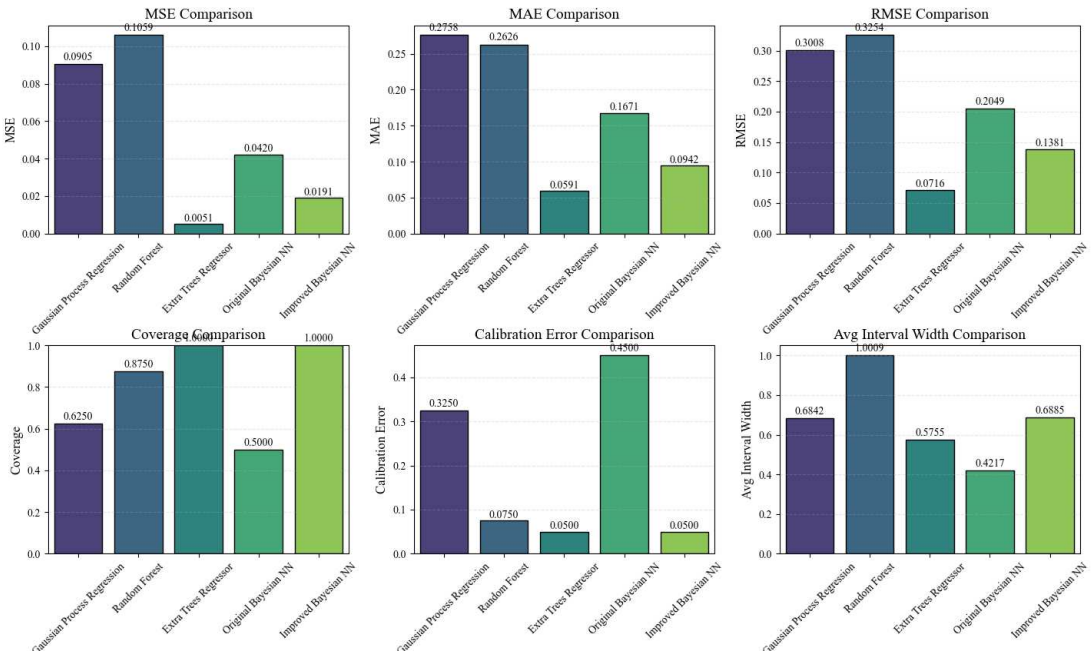


Fig.2. Comparison of model performance: quantification of accuracy and uncertainty

Regarding uncertainty quantification, for the Coverage indicator, both the Extra Trees Regressor and the Improved Bayesian NN achieved a value of 1.000, meaning their 95% confidence intervals could effectively cover the true values. The Coverage of Random Forest was 0.8750, that of Gaussian Process Regression was 0.6250, while the Coverage of the Original Bayesian NN was only 0.5000. For the Calibration Error indicator, both the Extra Trees Regressor and the Improved Bayesian NN had a calibration error of 0.0500, showing good performance. The calibration error of Random Forest was 0.0750, that of Gaussian Process Regression was 0.3250, and the calibration error of the Original Bayesian NN was as high as 0.4500. In terms of the Avg Interval Width, the Extra Trees Regressor had a relatively narrow interval width of 0.5755. The interval width of Random Forest was the largest, at 1.0009. The interval width of Gaussian Process Regression was 0.6842, that of the Original Bayesian NN was 0.4217, and the interval width of the Improved Bayesian NN was 0.6885. These indicators comprehensively reflect the basic performance of different models in terms of precision and uncertainty quantification.

3.2 Statistical significance verification of model differences

To verify the statistical significance of performance differences among models, pairwise t-tests were conducted, and the results are visualized through heatmaps, which is shown in Figure 3. These heatmaps illustrate the p-values from pairwise t-tests comparing five models (Gaussian Process Regression, Random Forest, Extra Trees Regressor, Original Bayesian NN, Improved Bayesian NN) in terms of accuracy (absolute error, left) and uncertainty (interval width, right). The color-coding of p-values follows the gradient bars, where values closer to 0 (darker blue) denote more significant differences.

For the accuracy comparison (assessed by absolute error, left heatmap), the p-values indicate that the Extra Trees Regressor shows significant differences ($p < 0.05$) when compared to the Gaussian Process Regression ($p = 0.0056$), Random Forest ($p = 0.0217$), and Original Bayesian NN ($p =$

0.0473). The Improved Bayesian NN exhibits significant differences only relative to the Gaussian Process Regression (p = 0.0043).

Regarding the uncertainty comparison (evaluated by interval width, right heatmap), the Original Bayesian NN demonstrates significant differences when contrasted with the Gaussian Process Regression (p = 0.0190) and Random Forest (p = 0.0170). The Improved Bayesian NN shows a significant difference from the Original Bayesian NN (p = 0.0288), signifying an enhanced ability in uncertainty quantification after improvement. Overall, these p-values statistically confirm the performance disparities among models, with the Extra Trees Regressor and Improved NN showcasing advantages in accuracy and uncertainty quantification, respectively.

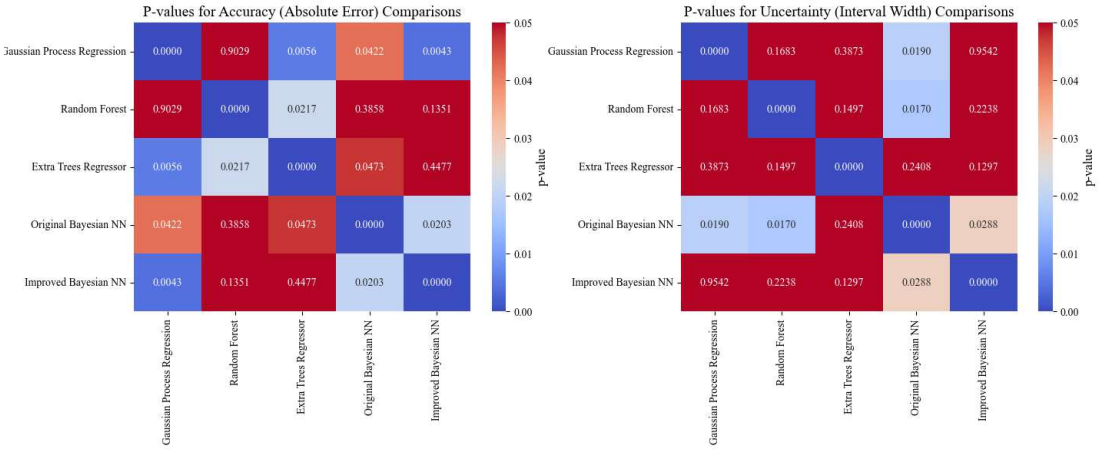


Fig.3. Pairwise t-test results for model performance differences

3.3 Feature sensitivity analysis

To analyze the response logic of models to input features, sensitivity analyses were conducted for eight features (CIC, CO_2, L, label, O_2, T, TP, W) using two models: the Extra Trees Regressor (ETR) and the Improved Bayesian Neural Network (ImprovedBNN). The results are shown in Figure 4-11. These paired sub-figures (one for each of the eight features: CIC, CO_2, L, label, O_2, T, TP, W) illustrate the response of two models (Extra Trees Regressor, ETR; Improved Bayesian Neural Network, ImprovedBNN) to feature fluctuations in terms of corrosion rate prediction and prediction uncertainty. The left sub-figure shows the "Change in corrosion rate prediction", with the blue line representing the ETR prediction, the red line representing the ImprovedBNN prediction, the purple - shaded area representing the ETR 95% confidence interval, and the pink-shaded area representing the BNN 95% confidence interval. The right sub-figure shows the "Uncertainty changes prediction", with the blue line representing the ETR standard deviation (STD) and the red line representing the ImprovedBNN STD.

For each feature, the change in corrosion rate prediction and the change in prediction uncertainty (represented by the standard deviation, STD) were evaluated. When analyzing the CIC feature, both the ETR and ImprovedBNN predictions of the corrosion rate showed a decreasing trend as the CIC fluctuation ratio increased, and the 95% confidence intervals of both models effectively covered the prediction changes. In terms of uncertainty, the STD of the ImprovedBNN remained relatively stable, while the ETR STD showed a slight decrease at higher fluctuation ratios.

For the L and W feature, the ImprovedBNN prediction of the corrosion rate exhibited significant fluctuations with changes in the L and W fluctuation ratio, whereas the ETR prediction remained stable. The uncertainty analysis also indicated that the ImprovedBNN STD had a prominent peak at the 0% fluctuation ratio, while the ETR STD was stable.

Overall, the ETR demonstrated stable prediction performance across most features, with relatively narrow and stable confidence intervals. The ImprovedBNN showed more significant responses to feature fluctuations, especially in terms of corrosion rate prediction changes, but also exhibited greater variability in uncertainty quantification. These results reveal the different response logics of the two models to input features: the ETR focuses on stable prediction with reliable uncertainty estimation, while the ImprovedBNN is more sensitive to feature changes but with more complex uncertainty behavior.

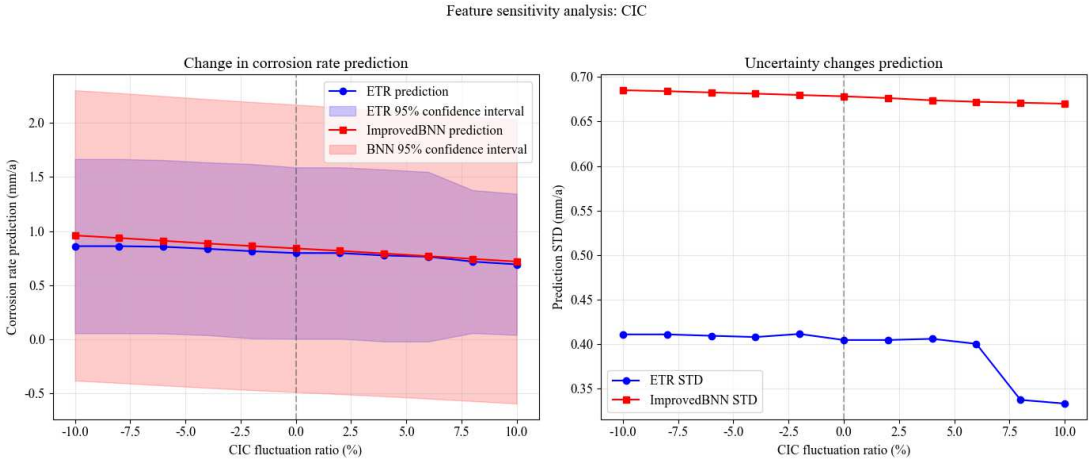


Fig.4. Feature sensitivity analysis for corrosion rate prediction models with CIC

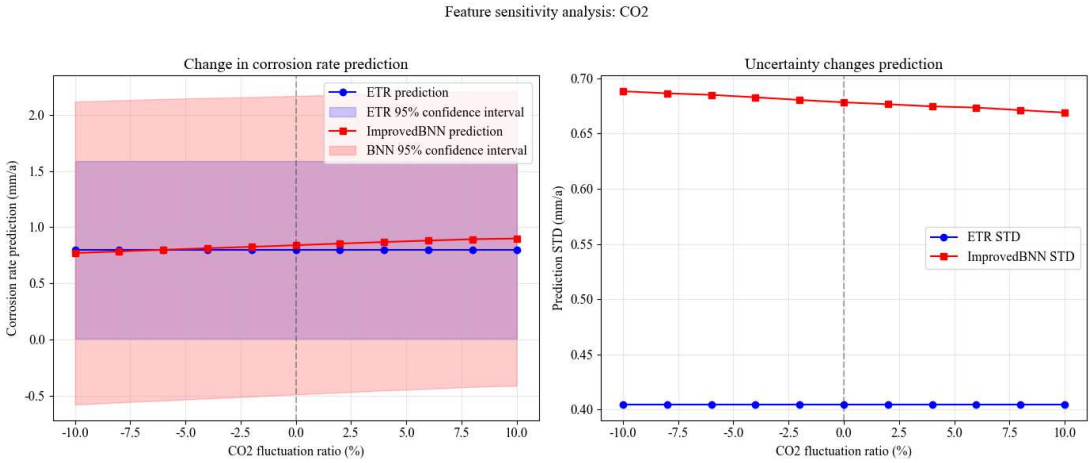


Fig.5. Feature sensitivity analysis for corrosion rate prediction models with CO_2

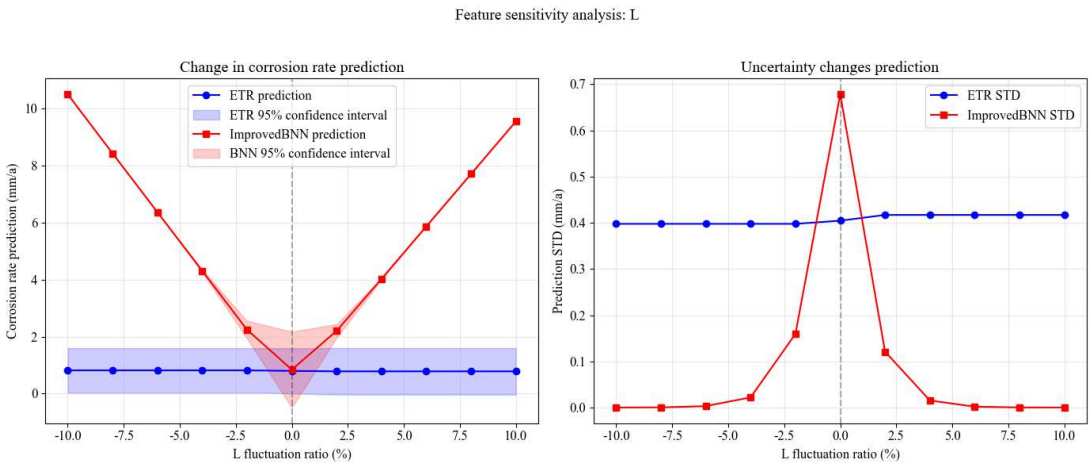


Fig.6. Feature sensitivity analysis for corrosion rate prediction models with L

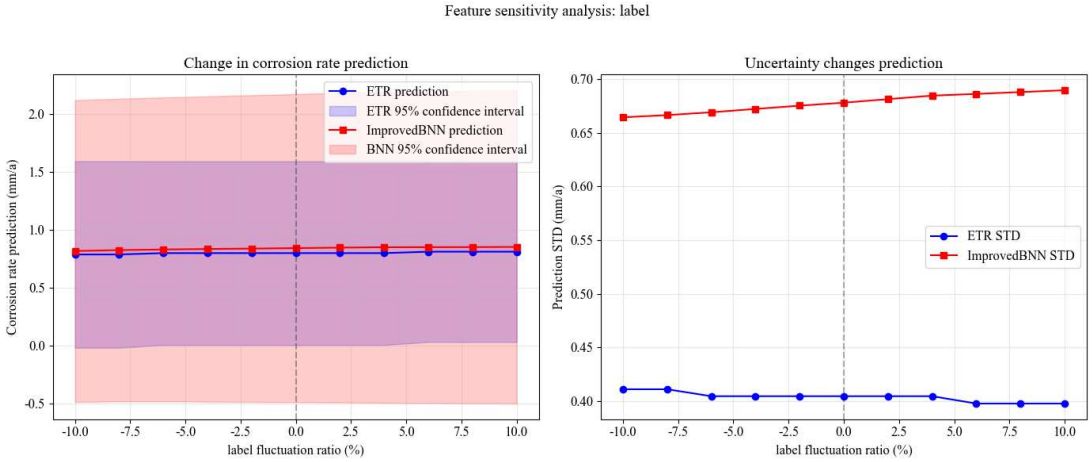


Fig.7. Feature sensitivity analysis for corrosion rate prediction models with label

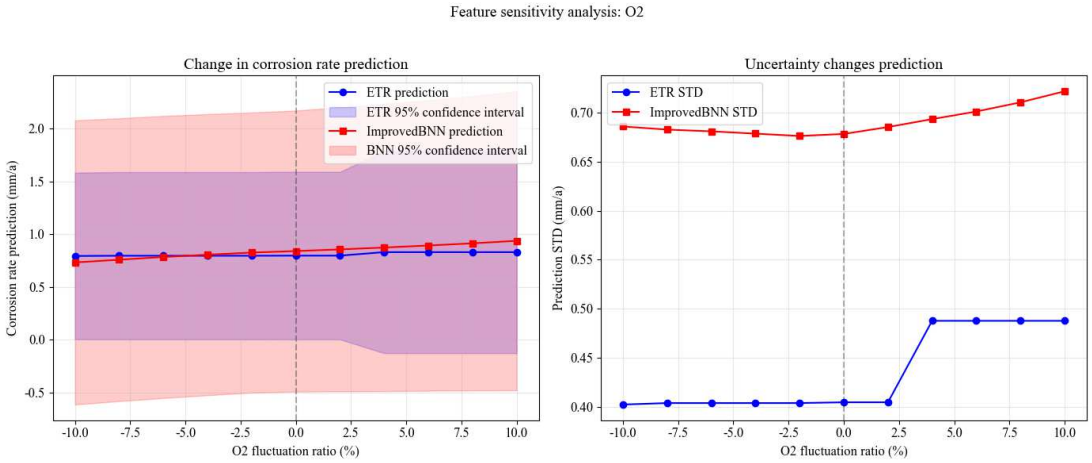


Fig.8. Feature sensitivity analysis for corrosion rate prediction models with O_2

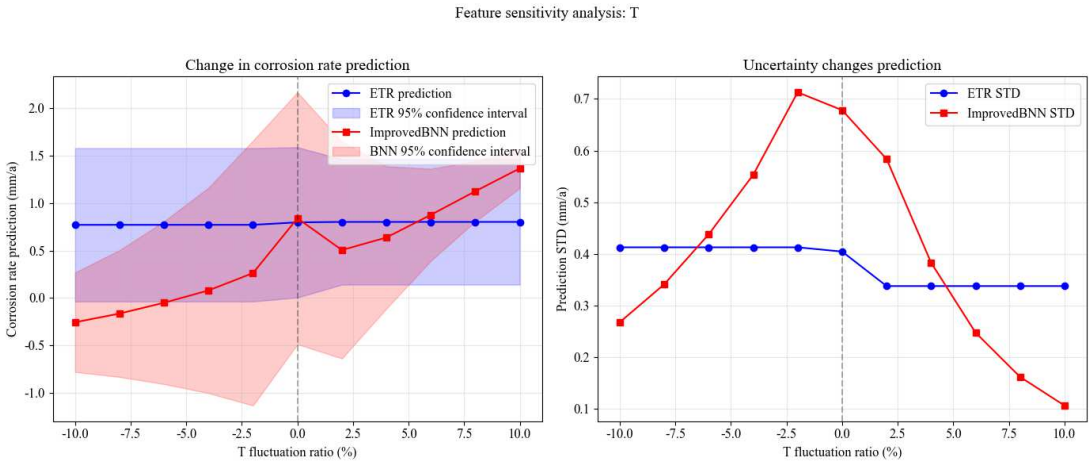


Fig.9. Feature sensitivity analysis for corrosion rate prediction models with T

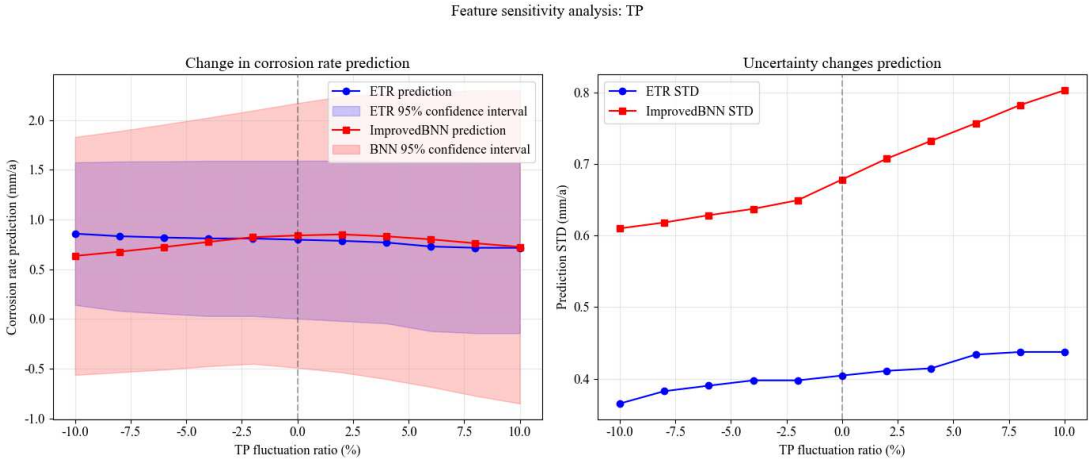


Fig.10. Feature sensitivity analysis for corrosion rate prediction models with TP

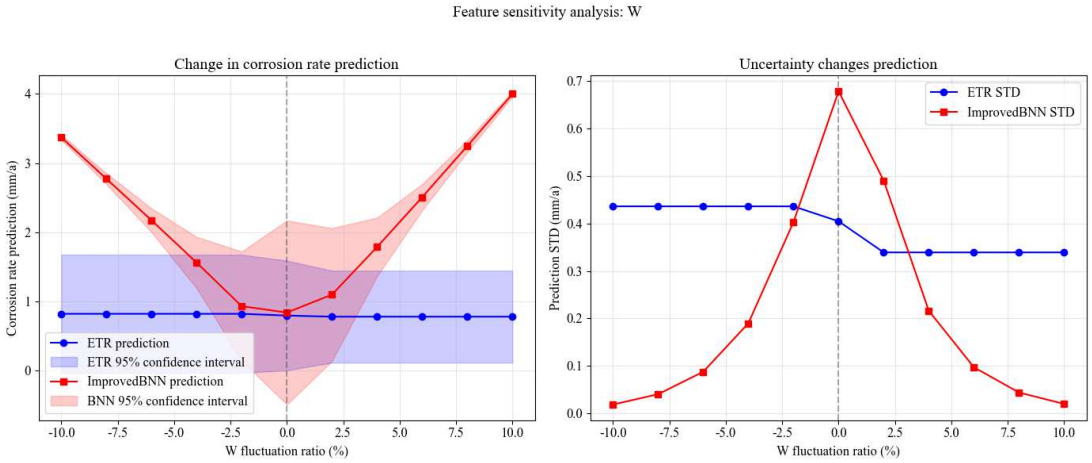


Fig.11. Feature sensitivity analysis for corrosion rate prediction models with W

Furthermore, the sensitivity ranking of different features is shown in Figure 12. The figure presents the normalized sensitivity of eight features (CIC, TP, W, O_2, L, T, label, CO_2) to corrosion rate predictions, as evaluated by two models: the Extra Trees Regressor (ETR, blue bars) and the Improved Bayesian Neural Network (ImprovedBNN, orange bars). Sensitivity values are

normalized such that higher bars indicate greater influence on prediction output.

For the ETR model, the CIC feature exhibits the highest sensitivity (normalized sensitivity close to 1.0), followed by TP (approximately 0.85). Other features, such as W, O_2, L, T, label, and CO_2, show relatively lower sensitivities, with CO_2 having the smallest value. CIC and TP directly regulate the electrochemical reaction rate of corrosion, so ETR (which excels at capturing strong nonlinear correlations) assigns them high weights. Furthermore, ETR (an ensemble tree model) focuses on features with high predictive power, leading to concentrated sensitivity on dominant factors (CIC/TP).

The Improved BNN demonstrates distinct behavioral characteristics compared to ETR, with the highest sensitivity to the structural length feature L (normalized sensitivity ≈ 1.0) and the second-highest to the width feature W (≈ 0.3), while CIC and TP have very low sensitivities. Notably, this sensitivity profile contrasts sharply with the ETR model, which prioritizes electrochemical features (CIC, TP) over structural dimensions. This divergence underscores the core trade-off between deterministic and probabilistic modeling. ETR focuses on features with strong predictive power for point estimates, while the Improved BNN prioritizes features that introduce uncertainty into corrosion evolution, thereby providing insights that are more directly relevant to risk-based corrosion management. Overall, the ETR and ImprovedBNN models display distinct sensitivity patterns across features, indicating different prioritization of feature impacts on corrosion rate prediction.

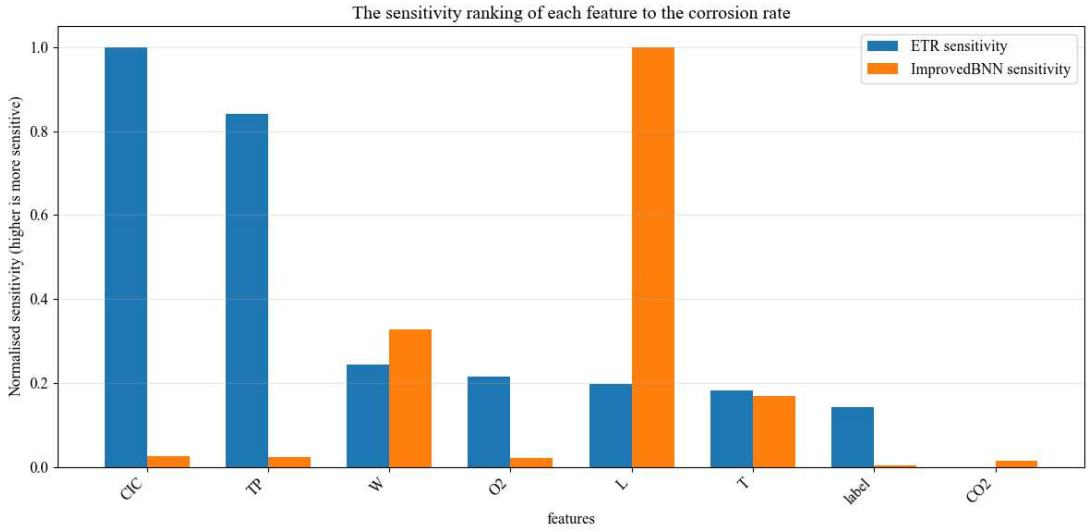


Fig.12. Normalized feature sensitivity to corrosion rate prediction across models

3.4 Ability to capture fluctuations: Robustness validation in dynamic scenarios

To verify the model's ability to capture preset fluctuations in dynamic scenarios, we adopted the controlled variable method to analyze the sensitivity of each feature to input changes. The controlled variable method refers to fixing all features except the target feature, and then analyzing the impact of the target feature on the model performance by observing the changes in the model's prediction results and uncertainty caused by the variation of the target feature within a certain range. When the input features change, the response tendency of the predicted value to feature fluctuations and the

uncertainty quantification of the adaptability to feature fluctuations are shown in Fig. 4-12. In Fig.4, these paired sub - figures (left: corrosion rate prediction change; right: uncertainty change) illustrate the models' response to CIC feature fluctuations (x-axis: CIC fluctuation ratio, from - 10% to + 10%).

Left sub-figure shows prediction response to feature fluctuations. The blue line (ETR prediction) and red line (ImprovedBNN prediction) show the trend of corrosion rate prediction changes as CIC fluctuates. Both models' predictions decrease slightly with increasing CIC fluctuation ratio, and the 95% confidence intervals (purple for ETR, pink for BNN) stably cover these changes, indicating consistent prediction responsiveness.

Right sub-figure shows uncertainty adaptation to feature fluctuations. The blue line (ETR STD) and red line (ImprovedBNN STD) represent prediction uncertainty. The ImprovedBNN STD remains stable (around 0.65–0.70), while the ETR STD shows a slight decrease at higher fluctuation ratios (dropping from ~ 0.42 to ~ 0.35). This reflects different uncertainty adaptation strategies: ETR adjusts uncertainty dynamically, while ImprovedBNN maintains stable uncertainty quantification.

4. Discussions

Structured around five key dimensions of model performance, the findings of this study yield critical insights into the behavioral characteristics of corrosion rate prediction models. Across precision metrics, the Extra Trees Regressor (ETR) consistently outperformed other models, exhibiting the lowest mean squared error (MSE = 0.0051), mean absolute error (MAE = 0.0591), and root mean squared error (RMSE = 0.0716) in point predictions. Meanwhile, the Improved Bayesian Neural Network (BNN) demonstrated marked enhancements over its original counterpart, reducing MSE from 0.0420 to 0.0191 and achieving perfect coverage (1.000) in uncertainty quantification, albeit with a narrower interval width (0.6885) than ETR. Statistical validation via paired t-tests confirmed these performance gaps. The prediction accuracy of ETR was significantly superior to Gaussian Process Regression, Random Forest, and the Original BNN ($p < 0.05$), while the uncertainty quantification of Improved BNN showed a decisive improvement over the Original BNN ($p = 0.0288$). Feature sensitivity analysis further highlighted model-specific response logics. ETR and the Improved BNN shared high sensitivity to critical features like CIC (normalized sensitivity > 0.8 for ETR), yet diverged in uncertainty adaptation-ETR dynamically adjusted interval widths, while the Improved BNN maintained stable confidence intervals.

At their core, these findings redefine the tradeoffs between deterministic and probabilistic modeling in corrosion science. The dominance of ETR in point prediction precision challenges the traditional preference for Bayesian frameworks in uncertainty-aware tasks, demonstrating that ensemble methods can achieve both accuracy and reliability when optimized for feature sensitivity. Conversely, the Improved BNN achieves balanced uncertainty quantification, outperforming its original counterpart and demonstrating competitiveness with ETR in terms of coverage, thus validating the significance of architectural refinements in neural networks for capturing complex nonlinear feature interactions. By resolving the tension between "precision" (ETR's strength) and "probabilistic rigor" (BNN's promise), this work establishes a new benchmark for model selection in corrosion rate forecasting. ETR emerges as the pragmatic choice for high-stakes, real-time

monitoring (e.g., pipeline integrity management), while the Improved BNN proves indispensable for scenarios demanding explicit uncertainty communication (e.g., risk-based inspection planning). In broader scientific contexts, this study bridges the gap between empirical model performance and theoretical uncertainty principles. It demonstrates that no single architectural design dominates universally; instead, the "optimal" model is contingent upon the specific requirements of corrosion engineering workflows, ranging from deterministic control to probabilistic risk mitigation.

4.1 Engineering application

From an engineering application perspective, the distinct performance characteristics of ETR and Improved BNN delineate clear applicability scenarios, directly translating research findings into practical corrosion management strategies.

For ETR, its outstanding precision in point predictions and stable, narrow confidence intervals make it highly suitable for scenarios where accurate deterministic forecasts are paramount. In pipeline integrity management, for example, real-time monitoring of corrosion rates requires models that can provide reliable, precise estimates to inform immediate maintenance decisions. ETR's ability to consistently deliver low-error predictions ensures that operators can trust the data to schedule inspections, replace components, or adjust operational parameters, minimizing the risk of unexpected failures. Additionally, in quality control processes for corrosion-prone materials, ETR can offer the accuracy needed to maintain product standards, reducing costs associated with rework or material waste.

On the other hand, the Improved BNN excels in probabilistic uncertainty quantification, thereby enabling its critical application in risk-based inspection planning and safety-critical engineering systems. In offshore oil and gas facilities, where corrosion risks are multifaceted and can have severe consequences, the explicit uncertainty communication of Improved BNN allows for a more nuanced assessment of risks. By providing not only a corrosion rate prediction, but also a measure of the uncertainty associated with that prediction, it enables engineers to prioritize inspection and mitigation efforts based on both the predicted corrosion rate and the level of confidence in that prediction. This probabilistic approach also supports better-informed decision-making in long-term asset management, helping to optimize resource allocation for corrosion prevention and control over the lifecycle of industrial infrastructure.

In summary, the engineering value of this research lies in its ability to guide the selection of the most appropriate corrosion rate prediction model for specific real-world needs. By aligning the distinctive strengths of the ETR and Improved BNN with specific engineering scenarios, this study equips corrosion engineers and asset managers to make more effective and targeted decisions, thereby ultimately enhancing the safety, reliability, and economic viability of corrosion-affected systems.

4.2 Limitation of the research methodology and mitigation

A key methodological limitation of this study is the restriction of feature sensitivity analysis to single-parameter fluctuations within typical operating ranges, with insufficient coverage of extreme conditions (e.g., supra-threshold corrosion environments or synergistic abrupt changes across

multiple features). This constraint implies that the generalizability of observed model response patterns remains unvalidated for complex, extreme feature combinations encountered in real industrial settings. Consequently, in unconventional scenarios, decisions on model selection and application derived from the current analysis may be suboptimal, potentially compromising the accuracy of extreme corrosion risk.

Nonetheless, the core conclusions of this work retain robustness and value. First, the focus on typical operating condition-encompassing fluctuation ranges commonly observed in industrial practice--ensures that the performance disparities and scenario-specific suitability of the Extra Trees Regressor (ETR) and Improved Bayesian Neural Network (ImprovedBNN) are rigorously validated. This provides a reliable foundation for model selection in most practical corrosion management contexts, such as routine pipeline monitoring and standard risk assessment. Second, the fundamental logic underlying the sensitivity analyses--specifically, the interplay between predictive accuracy, uncertainty quantification, and scenario adaptability-remains unaffected by the absence of extreme-condition data. The established framework linking "model performance to engineering utility" not only supports immediate application but also provides a methodological blueprint for extending analyses to extreme conditions.

This limitation further delineates clear avenues for future research, which can be advanced along three key directions. First, it is imperative to develop a comprehensive database for extreme operating conditions that incorporates synergistic multi-parameter perturbations. The database would enable rigorous validation of the model response boundaries under complex disturbance conditions. Second, it is necessary to design condition-adaptive predictive models that integrate the feature sensitivity framework identified in this study with custom-tailored compensation mechanisms for extreme conditions, thereby enhancing model robustness across the full operational spectrum. Third, the uncertainty quantification framework for extreme corrosion scenarios should be further advanced by building on the mechanistic insights into normal-condition uncertainty quantification derived in this study. Addressing these research gaps will further advance the engineering applicability of corrosion rate prediction models, facilitating their extension from typical operating scenarios to full operational coverage in industrial practice.

5. Conclusion

This study systematically evaluated the performance of multiple predictive models for corrosion rate forecasting, with a focus on comparing the Extra Trees Regressor (ETR) and an Improved Bayesian Neural Network (ImprovedBNN) across key dimensions: predictive accuracy, uncertainty quantification, feature sensitivity, and robustness to predefined fluctuations.

The results demonstrate that ETR achieves superior point prediction performance, with the lowest MSE (0.0051), MAE (0.0591), and RMSE (0.0716) among all models tested. It also exhibits reliable uncertainty quantification, with perfect coverage (1.000) and low calibration error (0.0500), supported by statistically significant differences ($p < 0.05$) compared to Gaussian Process Regression, Random Forest, and the Original Bayesian NN. The ImprovedBNN, while outperformed by ETR in point accuracy, shows marked advancements over its original counterpart, with reduced MSE (0.0191 vs. 0.0420) and enhanced uncertainty quantification (coverage = 1.000,

calibration error = 0.0500), establishing it as a robust probabilistic modeling alternative. Feature sensitivity analysis reveals distinct response logics. ETR prioritizes stable predictions with dynamically adjusted uncertainty intervals, while the ImprovedBNN exhibits higher sensitivity to feature fluctuations, making it suitable for scenarios requiring explicit uncertainty communication.

In practical applications, ETR is recommended for high-precision, real-time monitoring (e.g., pipeline integrity management), whereas the ImprovedBNN is better suited for risk-based decision-making (e.g., probabilistic inspection planning). Despite limitations in covering extreme multi-feature perturbation scenarios, the findings provide a validated framework for model selection in typical industrial settings and outline clear directions for future work, including the development of extreme-condition databases and adaptive models to enhance full-spectrum robustness.

Acknowledgements
The authors would like to express their sincere gratitude to Intelligent Geographical Telemetry and Information Perception Interdisciplinary Innovation Team for their financial support under grant 25WTJX02.

Conflict of interest
The authors declare that they have no conflicts of interest with respect to the research, authorship, and/or publication of this article.

Author Contributions
J. H. was responsible for the conceptualization, data collation, methodology design, and the preparation of the original draft as well as the review and editing of the manuscript. X. Z. participated in the acquisition of funding for the project, provided financial support for the research, and contributed to the review and editing of the manuscript. S. D. was engaged in the research investigation and made contributions to the progress of the research investigation process. Z. Y. undertook the supervision of the research, participated in data collection and project administration, oversaw the entire research process and the quality of the paper, and provided guidance on the overall direction and academic rigor of the paper.

Data Availability Statement
The data that support the findings of this study are available from the corresponding author upon reasonable request.

References
1. Wasim, M. & Djukic, M. B, (2022). External corrosion of oil and gas pipelines: A review of failure mechanisms and predictive preventions. J. Nat. Gas. Sci. Eng. 100, 104467.
2. Revie, R. W., & Uhlig, H. H. (2008). Corrosion and corrosion control: An introduction to corrosion science and engineering (4th ed.). John Wiley & Sons.
3. Chen, Z., Jin, Y., Wang, X., Chen, H., Zhu, B. & Li, W. (2025). Advancements in prediction models for corrosion in oil and gas pipelines. Corrosion Reviews. https://doi.org/10.1515/corrrev-2024-0119
4. Tafel, J. (1905). Über die Polarisation bei der Wasserstoffentwicklung. Zeitschrift für

Physikalische Chemie, 50, 641-681.
5. N. Anita, R.M. Joany, R. Dorothy, Jeenat Aslam, Susai Rajendran, A. Subramania, Gurmeet Singh, Chandrabhan Verma. (2023). Chapter 4 - Linear polarization resistance (LPR) technique for corrosion measurements, Editor(s): Jeenat Aslam, Chandrabhan Verma, Chaudhery Mustansar Hussain, Electrochemical and Analytical Techniques for Sustainable Corrosion Monitoring, Elsevier, 39(10), 491-496.
6. Nesic, S., Postlethwaite, J. (1994). Modelling of CO2 Corrosion Mechanisms. In: Trethewey, K.R., Roberge, P.R. (eds) Modelling Aqueous Corrosion. NATO ASI Series, vol 266. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-1176-8_15
7. ISO 12944-2:2017. *Paints and varnishes—Corrosion protection of steel structures by protective paint systems—Part 2: Classification of environments*. International Organization for Standardization.
8. Pourbaix, M. (1974). *Atlas of electrochemical equilibria in aqueous solutions*. Pergamon Press.
9. Popova, K.; Prošek, T. (2022). Corrosion Monitoring in Atmospheric Conditions: A Review. Metals, 12, 171.
10. Coelho, L.B., Zhang, D., Van Ingelgem, Y. *et al.* (2022). Reviewing machine learning of corrosion prediction in a data-oriented perspective. npj Mater Degrad 6, 8.
11. N. Daoudi, M. Y. Haouam, L. Laimeche, I. Bendib and M. Amroune, (2024). Applications of Machine Learning in Corrosion Detection, 2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS), EL OUED, Algeria, 1-8,
12. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
13. Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3-42.
14. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189-1232.
15. Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359-366.
16. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.
17. Yang, G., Li, Q., Wang, X., Zhu, R., Wang, L. (2025). Research on Data-Driven Time Series Prediction Models for Corrosion Current in Marine Atmosphere. In: Li, X. (eds) Proceedings of 22nd International Corrosion Congress. ICC 2024. Springer Proceedings in Materials, vol 71. Springer, Singapore.
18. Keren, Daniel & Werman, Michael. (1994). Bayesian Interpolation, Neural Computation.
19. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *International Conference on Machine Learning*, 1050-1059.
20. Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 5580–5590.
21. Terrados Cristos, Marta & Díaz-Piloñeta, Marina & Ortega-Fernández, Francisco & Martinez-Huerta, Gemma & Alvarez-Cabal, José. (2025). Corrosion Risk Assessment in Coastal Environments Using Machine Learning-Based Predictive Models. Sensors. 25. 4231. 10.3390/s25134231.
22. WANG Shasha, MA Shuaijie, CHE Kun, DU Yanxia. Application Status of Machine Learning

in Field of Natural Environment Corrosion Assessment and Prediction. Journal of Chinese Society for Corrosion and protection, 2023, 43(3): 441-451.

23. Yan L, Diao Y, Lang Z, Gao K. (2020). Corrosion rate prediction and influencing factors evaluation of low-alloy steels in marine atmosphere using machine learning approach. Sci Technol Adv Mater. 19;21(1):359-370.
24. Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95-International Conference on Neural Networks*, 4, 1942-1948.
25. Li, Q., Wang, X., Zhu, R., Yang, X. (2025). Study on Multi-factor Synergistic Corrosion of Q345 Steel in Marine Atmospheric Environment Based on Corrosion Big Data. In: Li, X. (eds) Proceedings of 22nd International Corrosion Congress. ICC 2024. Springer Proceedings in Materials, vol 71. Springer, Singapore.
26. Bishop, Christopher. (2006). Pattern Recognition and Machine Learning. 10.1117/1.2819119.
27. Blundell, C., et al. (2015). Weight uncertainty in neural networks. Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:1613-1622.
28. Lakshminarayanan, B., Pritchard, J., & Blundell, C. (2016). Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems, 30, 6402-6413.
29. Guo, C., et al. (2017). On calibration of modern neural networks. International Conference on Machine Learning, 1321-1330.
30. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.
31. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 4768–4777.
32. Yin, Z.F., Yu, S.Z., Ma, B. et al., (2024). Corrosion Evaluation and Mechanical Life Prediction of J55 Tubing Steel under Air-Foam Flooding Oil Well Conditions. J. of Materi Eng and Perform. https://doi.org/10.1007/s11665-024-09780-1
33. Hu, J.L. (2024). Prediction of the internal corrosion rate for oil and gas pipelines and influence factor analysis with interpretable ensemble learning. Int. J. Press. Vessels Pip. 212, 105329.