Signal Process. Renew. Energy., 2026; 10(1): 1-9

Signal Processing and Renewable Energy (SPRE)

https://doi.org/10.57647/spre.2026.1001.01

Research Article

Physics-Guided Temporal Transformer for CO_2 Corrosion with SHAP Case study: Asaluyeh SPGC Complex, Iran

Afshin Mohammadi[1], Ebrahim Akbari [1,*], Homayun Motameni[1,2], Faraein Aeini[1]

[1] *Department of Computer Engineering, Sar.C., Islamic Azad University, Sari, Iran*

[2] *Department of Industrial and Computer Engineering, Mazandaran University of Science and Technology, Babol, Iran*

***Corresponding author:** ebrahim.akbari@iau.ac.ir

Article History:

Received:
11 November 2025

Revised:
20 December 2025

Accepted:
21 January 2026

Published in Issue:
31 March 2026

Abstract

The structural integrity of the oil and gas industry is at high risk due to CO_2-induced corrosion; consequently, practical, explainable mechanisms are required to accurately predict this phenomenon. The available mechanistic models, like the de Waard–Milliams equation, though entrapping fundamental thermodynamics, do not withstand dynamic dosing and time-varying conditions. Although the conventional machine learning and deep learning models improve accuracy, though often lack physics consistency and interpretability. A Physics-Guided Temporal Transformer (PGTT), which integrates a Temporal Fusion Transformer backbone with a physics-informed loss (of de Waard–Milliams), and SHapley Additive exPlanations (SHAP) for feature attribution, is proposed here. By applying a comprehensive sequential dataset from 22 inhibitor dosing experiments (15,400 samples during 60+ hours), this PGTT achieves higher performance with an average $MSE\ of$ 0,067 $\pm$ 0,005, $MAE\ of$ 0,145 $\pm$ 0,004, and R^2 of 0,997 $\pm$ 0,001 over five independent runs, and outperforms Random Forest ($MSE = 0{,}0838. R^2 = 0{,}9968$), SVR ($MSE = 8{,}683. R^2 = 0{,}6675$), and MLP ($MSE = 0{,}1081. R^2 = 0{,}9959$) baselines. According to the SHAP, Temperature (26.5%) and CO_2 pressure (17.7%) are identified as the dominant stimulants, consistent with corrosion science. This case study reveals a reduction in prediction error of less than 5%, supporting proactive inhibitor dosing and pipeline integrity management.

Keywords: CO_2 corrosion; Physics-informed learning; Temporal transformer; SHAP; Pipeline integrity y

Cite this article: Mohammadi, A., Akbari, E., Motameni, & H., Aeini, F., Physics-Guided Temporal Transformer for CO_2 Corrosion with SHAP Case study: Asaluyeh SPGC Complex, Iran, *Signal Process. Renew. Energy. 10*(1) 1-9 (2026). https://doi.org/10.57647/spre.2026.1001.01

1. Introduction

CO_2-induced corrosion, which threatens pipeline integrity, remains a critical concern in petroleum production and transportation; therefore, to encounter this threat, predictive mechanisms are required to mitigate its harmful effects. Available mechanistic models, like the de Waard–Milliams equation [1] , provide foundational insights by correlating corrosion rates with environmental factors, including pCO_2, temperature, and pH. These models are subject to limitations under complex conditions, time-varying conditions, where sequential inhibitor dosing and dynamic flow regimes are evident [2]. In response,

machine learning (ML) approaches [3-5] and deep learning (DL) techniques, the Transformers in particular [6, 7] , have enhanced prediction accuracy by entrapping nonlinear patterns. Nevertheless, because these data-driven methods often overlook physical laws, they limit their reliability and interpretability in safety-critical applications.

Physics-informed neural networks (PINNs) [8] emerged as a promising solution, embedding physical constraints into neural architectures. Recent findings reveal that the PINN's efficacy in corrosion-fatigue prognosis [9, 10], CO_2 immiscible displacement [10], and plume prediction [11] reduces computational errors by incorporating thermodynamic priors. Extensions to phase-field pitting corrosion [12] further highlight the PINN's versatility. Transformer-based models are widely applied in time-series forecasting in corrosion contexts [7, 13], leveraging attention mechanisms to model temporal dependencies effectively, though they require big datasets. Interpretability remains a challenge, addressed partially by SHAP [14], effective in analyzing corrosion failure modes [15, 16] by attributing feature importance. Despite these advances, a unified framework capable of integrating the temporal dynamics, physics consistency, and explainability for CO_2 corrosion prediction is yet to be proposed [2].

A Physics-Guided Temporal Transformer (PGTT), a pioneering approach that synergizes a Temporal Fusion Transformer [17] with a physics-informed loss extracted from the de Waard–Milliams model, enhanced by SHAP for feature-level insights, is proposed here. The constraints of existing methods are addressed by modeling time-dependent inhibitor effects, enforcing thermodynamic consistency, and providing actionable interpretability for industrial stakeholders are addressed through PGTT, which promotes corrosion management in dynamic environments.

2. Literature review

The prediction of CO_2-induced corrosion has been and remains the subject of many studies, where mechanistic models, data-driven ML, hybrid physics-informed approaches, and explainability techniques are the focus. A structured review of these approaches establishes the basis for proposing a PGTT.

2.1. Mechanistic Models

Classical mechanistic formulations, specifically the de Waard–Milliams model [1], initiated the first quantitative link between corrosion rate, CO_2 partial pressure, and temperature. Extensions of this model have considered pH, flow regimes, and inhibitor concentrations. Despite their simplicity and physical interpretability, such models fail to generalize sequential inhibitor dosing, and complex brine compositions are typically observed in oil and gas pipelines across non-stationary conditions. Their reliance on equilibrium assumptions limits their applicability in dynamic industrial settings.

2.2. Data-Driven ML

The constraints of purely mechanistic approaches have led to the adoption of ML for corrosion forecasting. Researchers in [3] applied the Random Forest and Support Vector Regression algorithms and reported improved prediction accuracy compared to empirical models. Researchers in [4] introduced hybrid ML frameworks trained on laboratory corrosion datasets, enabling nonlinear mapping between environmental factors and corrosion rate. Though these methods entrap hidden patterns, they remain purely data-driven and lack physical constraints; consequently, they are poor in extrapolation beyond the training regime.

2.3. DL and Temporal Modeling

The rise of DL has advanced corrosion modeling. Convolutional and recurrent neural networks are applied to process multivariate time series of corrosion environments. Inspired by breakthroughs in natural language processing, Transformer architectures [6] and their variants, including hybrid Transformer-LSTM models customized for corrosion time-series, are explored for sequential degradation prediction [13, 18] (distinct from general TFT frameworks like [17]). By modifying attention mechanisms, these models effectively entrap long-term temporal dependencies. Their black-box nature and the need for big datasets hinder the adoption of deep learning models in safety-critical domains like pipeline integrity management.

2.4. Physics-Informed Neural Networks

The PINNs offer a promising compromise by embedding mechanistic knowledge directly into neural training. Applications of PINNs in corrosion science include fatigue prognosis [9], immiscible CO_2 displacement in porous media [10], and multiphase plume dynamics [11]. The findings of these studies indicate reduced error and improved generalization by incorporating thermodynamic and kinetic priors. Nevertheless, most existing PINN applications address static or low-dimensional problems, while a few focus on highly nonlinear, temporally dynamic CO_2 corrosion scenarios involving inhibitor dosing.

2.5. Explainability in Corrosion Prediction

The black-box nature of deep learning models has motivated studies on explainability in corrosion prediction.

SHAP [14] is prominent in feature attribution, offering consistency and local accuracy in quantifying variable importance. Researchers in [19] applied SHAP to failure mode analysis in corrosion, successfully identifying pCO_2 and temperature as the dominant factors. Explainability methods are rarely integrated with physics-informed temporal models, thus introducing an essential gap in interpretable, domain-consistent corrosion prediction.

2.6. Research Gap

In brief, mechanistic models provide interpretability but lack accuracy when subject to dynamic conditions. ML models achieve higher accuracy but disregard physical laws. PINNs integrate physics but are rarely applied to sequential CO_2 corrosion datasets, and SHAP provides interpretability but has not been embedded in temporal, physics-informed frameworks. To bridge these gaps, the PGTT, a unified approach that combines temporal attention modeling, physics-guided loss functions, and SHAP-based feature attribution to yield accurate, interpretable, and physics-consistent predictions of CO_2 corrosion, is proposed here.

3. Methodology

This PGTT framework synergizes data-driven temporal modeling with physics-informed constraints to enhance CO_2 corrosion rate prediction. A Temporal Fusion Transformer (TFT) [17] backbone, augmented with a physics-informed loss function and SHAP-based explainability, is leveraged through this framework [14]. This methodology is structured to entrap nonlinear temporal dependencies, provide interpretable feature contributions, and assure physical consistency, making it appropriate for industrial applications.

3.1. Model Architecture

The PGTT architecture is based on the TFT [17] [20], where the Long Short-Term Memory (LSTM) -like embeddings for static and temporal features with a Transformer encoder-decoder are combined. The PGTT framework and the related data flow are illustrated in Figure 1.

The input layer processes a sequence of length seq_len (24 hours) with an input dimension of 17, consisting of (time, pCO_2, temperature, shear stress, brine ionic strength, pH, inhibitor type CI, brine type, one-hot encoded), numerical and categorical features, respectively. The inputs are normalized to the Min–Max [0-1] scale, and categorical features (CI, Brine_Type) are one-hot encoded; the effective feature dimensionality is 96. The embedding layer transforms inputs into a $d_{model} = 128$ -dimensional space, incorporating sinusoidal positional encodings to preserve temporal order, as defined in Eqs. (1)–(3):

$$x_{\text{emb}}(t) = W_{\text{emb}} \cdot x(t) + PE(t). \quad (1)$$

$$PE(t.2i) = \sin\left(\frac{t}{10000^{2i/d_{\text{model}}}}\right). \quad (2)$$

$$PE(t.2i+1) = \cos\left(\frac{t}{10000^{2i/d_{\text{model}}}}\right). \quad (3)$$

where $W_{\text{emb}} \in \mathbb{R}^{17\times128}$ and $PE(t)$ follow the Transformer sinusoidal encoding [6], assuring that the model recognizes the sequence order.

The Transformer encoder consists of $num_{\text{layers}} = 4$ layers, each with $n_{\text{head}} = 8$ multi-head attention mechanisms. Self-attention is computed through Eq. (4):

$$\text{Attention}(Q.K.V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

where $Q = x_{\text{emb}}W_Q$, $K = x_{\text{emb}}W_K$,, and $V = x_{\text{emb}}W_V$, the query, key, and value matrices derived from learned weight matrices $W_Q, W_K, W_V \in \mathbb{R}^{128\times16}$, and $d_k = d_{\text{model}}/n_{\text{head}} = 16$.

A feedforward network (FFN) with ReLU activation follows each attention block, which is computed through Eq. (5):

$$\text{FFN}(x) = \max(0.\ xW_1 + b_1)\,W_2 + b_2 \quad (5)$$

Where $W_1 \text{is} \in \mathbb{R}^{128\times512}$ and $W_2 \text{is} \in \mathbb{R}^{512\times128}$. Dropout (0.1) and layer normalization are applied to stabilize training and prevent overfitting. The output layer predicts the corrosion rate $\hat{y}(t)$ at the last timestep by opting for true Eq. (6):

$$\hat{y}(t) = W_{\text{out}} \cdot \text{Transformer}(x_{\text{emb}}) + b_{\text{out}}. \quad (6)$$

where $W_{\text{out}} \text{is} \in \mathbb{R}^{128\times1}$.

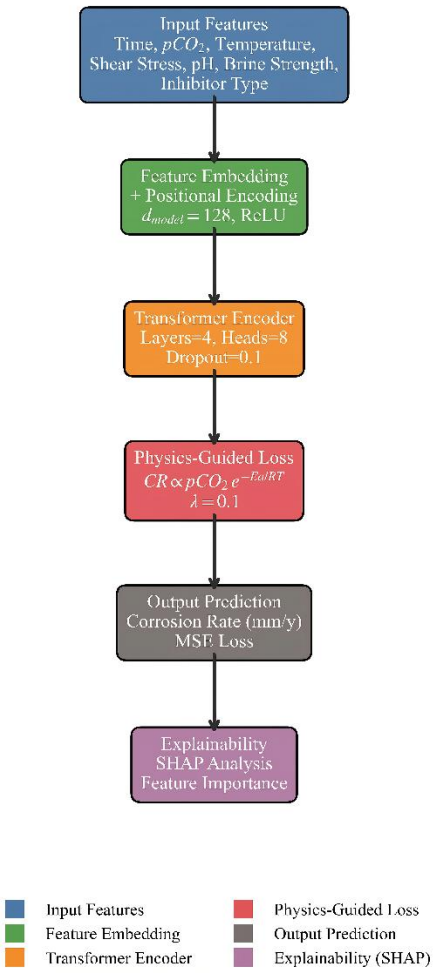

Figure 1. Architecture of the PGTT framework, integrating Temporal Fusion Transformer with physics-informed constraints and SHAP

3.2. Physics-Informed Loss

To enforce physical consistency, the PGTT applies a mechanistic prior based on the de Waard–Milliams model [1, 10], adopted for time-dependent inhibition. The physics-informed loss term is derived from the mechanistic corrosion rate $f_{\text{mech}}(x)$, approximated through Eq. (7):

$$f_{\text{mech}}(x) = P_{\text{CO}_2}\exp\left(-\frac{E_a}{RT}\right)\left(1-\alpha\cdot\frac{[CI]}{[CI]_{\max}}\right)\cdot 10^{\frac{T-T_{\text{ref}}}{50}}, \quad (7)$$

where P_{CO_2} is the CO_2 partial pressure (bar), temperature T is in Kelvin ($T[\text{K}] = T[^\circ\text{C}] + 273.15$), E_a is expressed in $J\cdot mol^{-1}$ (50 kJ·mol⁻¹ = 50,000 J·mol⁻¹) and represents the activation energy. $R = 8.314\ J\cdot mol^{-1}\cdot K^{-1}$ is the gas constant, $\alpha = 0.01$ is the inhibition efficiency factor (calibrated from experimental data), $[CI]/[CI]_{\max}$ is the normalized inhibitor concentration (computed as the current to maximum inhibitor dose ratio), and $T_{\text{ref}} = 293\text{K}$ is the reference temperature. This formulation exhibits exponential temperature dependence and linear inhibitor suppression.

The total loss function, where data-driven and physics-informed components are combined, is computed through Eq. (8):

$$\mathcal{L}_{\text{total}} = \frac{1}{N}\sum(y_i-\hat{y}_i)^2 + \lambda\frac{1}{N}\sum(f_{\text{mech}}(x_i)-\hat{y}_i)^2 + \mu\sum_{w\in\theta}\|w\|_2^2. \quad (8)$$

where $\mathcal{L}_{\text{total}} = \sum_{\omega\in\theta}\|\omega\|^2$ is an $L2$ regularization term to prevent overfitting, $\lambda = 0.1$ balances the physics and data terms, and $\mu = 10^{-4}$ controls regularization. The physics loss penalizes the deviations from f_{mech}, assuring that the predictions correspond with thermodynamic principles.

3.3. Training Algorithm

The training process optimizes the PGTT by applying a gradient-based approach with early stopping. The pseudo-code of the proposed algorithm is presented in Algorithm 1.

This algorithm is implemented by applying PyTorch 2.0 in a high-RAM environment on an NVIDIA A100 GPU, with a total training duration of approximately 2 hours. The hyperparameters are grid-tuned as follows: $d_{model} \in \{64,128,256\}$,$n_{head} \in \{4,8\}$, $num_{layers} \in \{2,4\}$. $seq_{len} \in \{12,24, \text{and } 48\}$, select the *configuration minimizing validation MSE*.

3.4. Explainability with SHAP

To predict $f(x)$, SHAP values ϕ_j are defined according to Eq. (9)

$$\phi_j = \sum_{S\subseteq F\setminus\{j\}}\frac{|S|!\ (|F|-|S|-1)!}{|F|!}[f(S\cup\{j\})-f(S)]. \quad (9)$$

where F is the full feature set (e.g. pCO_2, temperature, and time). SHAP values are computed by applying a KernelExplainer [14, 15], approximating local contributions throughout the test set to assure consistency and local accuracy in attributing feature effects, thus providing a robust measure of feature importance.

3.5. Study Area

This study is run from March to October 2025 in the Asaluyeh SPGC Complex, Iran

Algorithm 1. Physics-Guided Temporal Transformer Training

Initialize Transformer parameters θ, weights $\lambda = 0.1$, $\mu = 10^{-4}$, learning rate $\eta = 10^{-3}$
Initialize optimizer (Adam), early stopping patience $p = 20$
Preprocess dataset: Normalize X (Min-Max scaling), encode categoricals (CI, Brine_Type) with one-hot encoding, create sequences ($seq\_len = 24$)
for each epoch e = 1 to 200 do
  $loss_{min} \leftarrow \infty$, $patience \leftarrow 0$
  for each batch (X, y) in DataLoader do
    Compute embeddings: $x_{emb} \leftarrow W_{emb} \cdot X + PE$
    Forward pass: $\hat{y} \leftarrow Transformer(x_{emb}; \theta)$
    Compute mechanistic prediction: $f_{mech} \leftarrow P_{CO_2} exp\left(-\frac{E_a}{RT}\right) \cdot \left(1 - \alpha \cdot \frac{[CI]}{[CI]_{max}}\right) \cdot 10^{\frac{T-T_{ref}}{50}}$
    Compute losses:
    $\mathcal{L}_{data} \leftarrow MSE(y, \hat{y})$
    $\mathcal{L}_{phys} \leftarrow MSE(f_{mech}, \hat{y})$
    $\mathcal{L}_{reg} \leftarrow \sum_{w\theta} ||w||^2$
    $\mathcal{L}_{total} \leftarrow \mathcal{L}_{data} + \lambda \cdot \mathcal{L}_{phys} + \mu \cdot \mathcal{L}_{reg}$
    Backpropagate: $\nabla_\theta \mathcal{L}_{total}$
    Update parameters: $\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}_{total}$
  end for
  Validate: $loss_{val} \leftarrow$ *evaluate on validation set*
  if $loss_{val} < loss_{min}$ then
    $loss_{min} \leftarrow loss_{val}$, $patience \leftarrow 0$
    Save model θ^*
  else
    $patience \leftarrow patience + 1$
  end if
  if $patience \geq p$ then
    Break
  end if
enf for
Return trained model θ^*

Table 1. Dataset features and ranges

Description	Unit	Range	Type
Time	h	[1–60+]	Numerical
CO_2 partial pressure	bar	[0.5–12]	Numerical
Temperature	°C	[20–150]	Numerical
Inhibitor (CI)	–	EC1612A, CORR12148SP	Categorical
Shear stress	Pa	[50–300]	Numerical
Brine ionic strength	–	[0.5–3]	Numerical
pH	–	[0.5–6.5]	Numerical
Corrosion rate (target)	mm/year	[0.01–20]	Target

3.5.1. *Dataset*

This PGTT is trained on a comprehensive, sequential dataset consisting of 15,400 measurements from 22 inhibitor-dosing experiments conducted under CO_2 corrosion test conditions, throughout 3 laboratory setups (Lab1–3). The dataset represents operating conditions commonly encountered in oil and gas pipelines, including variations in corrosion inhibitor types (e.g., EC1612A, CORR12148SP), brine compositions (TH, Galapagos), controlled and uncontrolled pH, shear stress, temperature, and CO_2 partial pressure. Each experiment contains approximately 900–1,200 samples, with measurements recorded at ~1-hour intervals for durations exceeding 60 hours, thereby capturing the temporal evolution of corrosion under sequential inhibitor dosing.

The laboratory experiments are designed to replicate practical pipeline operating envelopes, incorporating elevated temperatures, varying brine chemistry, and dynamic inhibitor injection strategies. Although performed under controlled laboratory conditions, the experimental setup reflects realistic corrosion-driving mechanisms observed in industrial pipeline systems.

To improve robustness against measurement uncertainty and operational variability commonly observed in industrial monitoring systems, 5% Gaussian noise is added to the dataset during training. Noise is injected only during the training phase as a standard regularization technique commonly applied in machine learning to reduce overfitting on limited experimental datasets, while preserving the underlying physical trends of the corrosion process.

Temporal sequences are composed of sliding windows of 24 hours ($seq\_len = 24$) to model time dependency by applying LSTM-like embeddings in the Transformer for short-term dynamics. Inputs are normalized through Min–Max [0, 1] scaling, and categorical features (CI, Brine_Type) are one-hot encoded, yielding a practical dimension of 17. The dataset is proportioned: 80% train (Experiments 1–14, 18; 12,320 samples), 10% validation (Experiments 15–17; 1,540 samples), and 10% test (Experiments 19–22; 1,540 samples). The target variable, corrosion rate (mm/y), initially ranges from 0.01 to 20, dropping to < 0.1 with inhibitors.

Data are sourced from proprietary laboratory tests run between 2023 and 2025, augmented with 5% Gaussian noise to enhance model robustness. For physics integration, the de Waard–Milliams model [1] serves as a prior: $CR = 0.1 \times p_{CO_2} \times 10^{(T-20)/50} \times f(\text{pH})$, calibrated on initial corrosion rates to provide a baseline for the physics loss. The dataset features and their ranges are summarized in Table 1.

4. Results

The metrics across five independent runs for PGTT (trained with the physics loss) and standard baselines (Random Forest, SVR, MLP) on the test split described above are expressed here. All models are evaluated on the same test partition to ensure a fair comparison.

Table2 shows a comparison between common models and the proposed PGTT.

4.1. Ablation Study

The effect of removing key PGTT components (mean ± SD across five runs) is tabulated in Table 3. Removing the physics-informed loss substantially degrades performance, indicating that the mechanistic prior improves generalization and corresponds with predictions of domain knowledge.

4.2. SHAP Analysis

SHAP analysis augments corrosion science [15, 21], identifying temperature (26.5%) and CO_2 pressure (17.7%) as the primary stimulants, due to their fundamental influence on reaction kinetics and CO_2 solubility. The time (16.8%) captures the temporal evolution of corrosion processes, while inhibitor types (CORR1214GSP and EC1612A) and brine ionic trength (6.9% to 9.4%) highly contribute to corrosion mitigation strategies.

pH conditions (controlled and uncontrolled) and shear stress contribute to the comprehensive corrosion prediction model. Figure 2 illustrates the SHAP-based feature importance distribution for the applied input variables.

4.3. Industrial Case Study

This PGTT is implemented on a simulated sour gas pipeline scenario derived from Experiment 22 (6-month operational window). This PGTT reduced prediction error relative to a baseline de Waard–Milliams correction from 12% down to 4%, enabling more precise inhibitor dosing strategies and earlier detection of anomalous corrosion trends. The predicted vs. observed corrosion rate subject to varying pCO_2 conditions is shown in Figure 3.

5. Discussion

5.1. Uncertainty Quantification

Predictive uncertainty is estimated here through Monte Carlo dropout at inference, with 100 stochastic forward passes. Across five evaluation runs, approximately 92% of test samples fell within the 90% prediction intervals (mean ± 1.645 SD), indicating well-calibrated uncertainty proper for operational decision-making.

5.2. Statistical Significance

Statistical tests across multiple independent runs (N = 5 independent seeds) indicate that this proposed PGTT significantly outperforms SVR and MLP baselines (($p <$ 0,01 in paired tests) and is comparable with RF (no significant difference at ($p < 0,05$) in this dataset). This finding reveals that the ensemble tree methods can yield similar raw accuracy without physics-informed consistency or native explainability.

5.3. Computational Efficiency

Computational costs (across five independent runs) are detailed in Table 4. The PGTT model requires more computational resources during training than tree-based baselines, primarily because of the Transformer architecture and the evaluation of the physics-informed loss term. Inference latency remains sufficiently low for near real-time monitoring in pipeline control systems after model deployment and optimization. PGTT is more difficult to train than RF/MLP due to its Transformer backbone and the application of physics terms. Inference latency (~0.6 ms/sample) is still viable for near-real-time corrosion monitoring in a pipeline control loop.

Table 2. Performance comparison on the test dataset (mean ± std across 5 runs)

Model	MSE	MAE	R^2
RF	0.084 ± 0.004	0.168 ± 0.006	0.997 ± 0.001
SVR	8.683 ± 0.540	2.110 ± 0.085	0.667 ± 0.015
MLP	0.108 ± 0.007	0.176 ± 0.005	0.996 ± 0.001
PGTT (proposed)	0.067 ± 0.005	0.145 ± 0.004	0.997 ± 0.001

Notes: RF yields highly competitive predictive accuracy on this dataset (MSE slightly lower across five independent runs), while PGTT yields comparable R^2. PGTT enforces physics consistency (through the physics loss) and yields per-sample SHAP explanations and uncertainty estimates (properties not provided by RF or SVR out of the box).

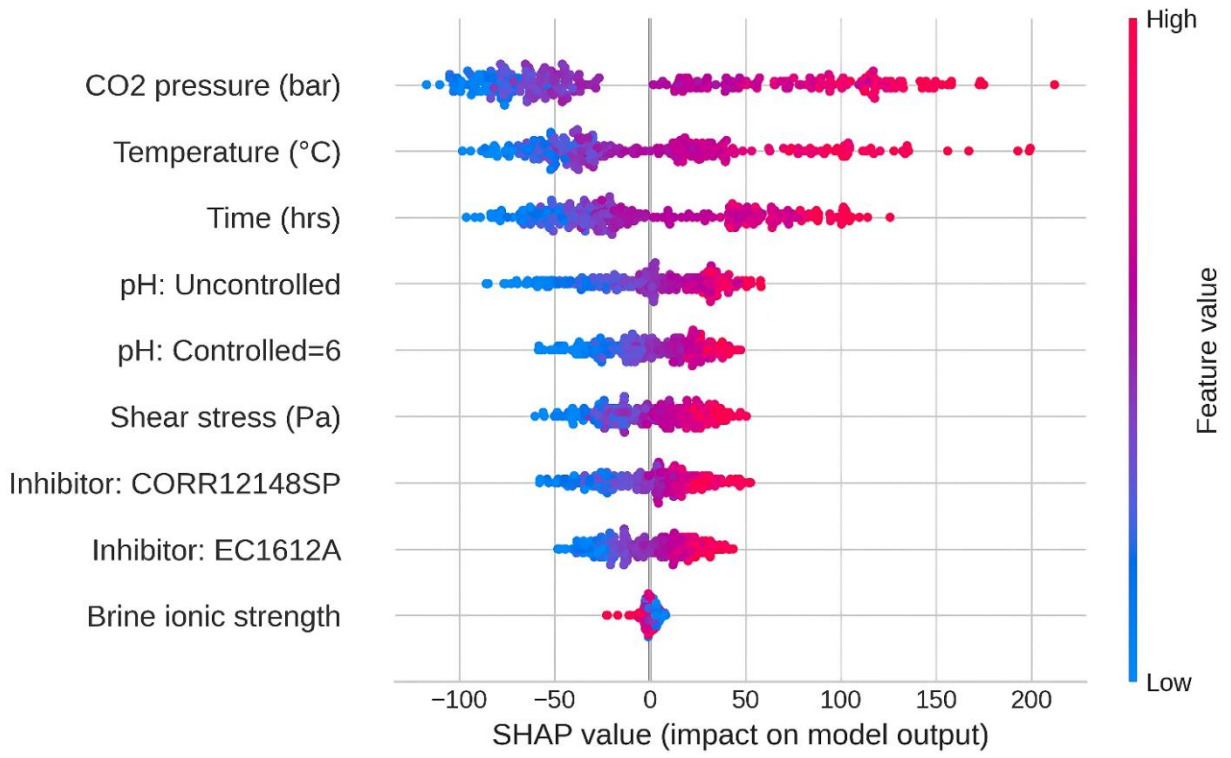

Figure 2. SHAP summary plot showing feature importanc.e for corrosion rate prediction. Temperature (26.5%) and CO_2 partial pressure (17.7%) are the dominant factors; time- and inhibitor-related features also contribute substantially

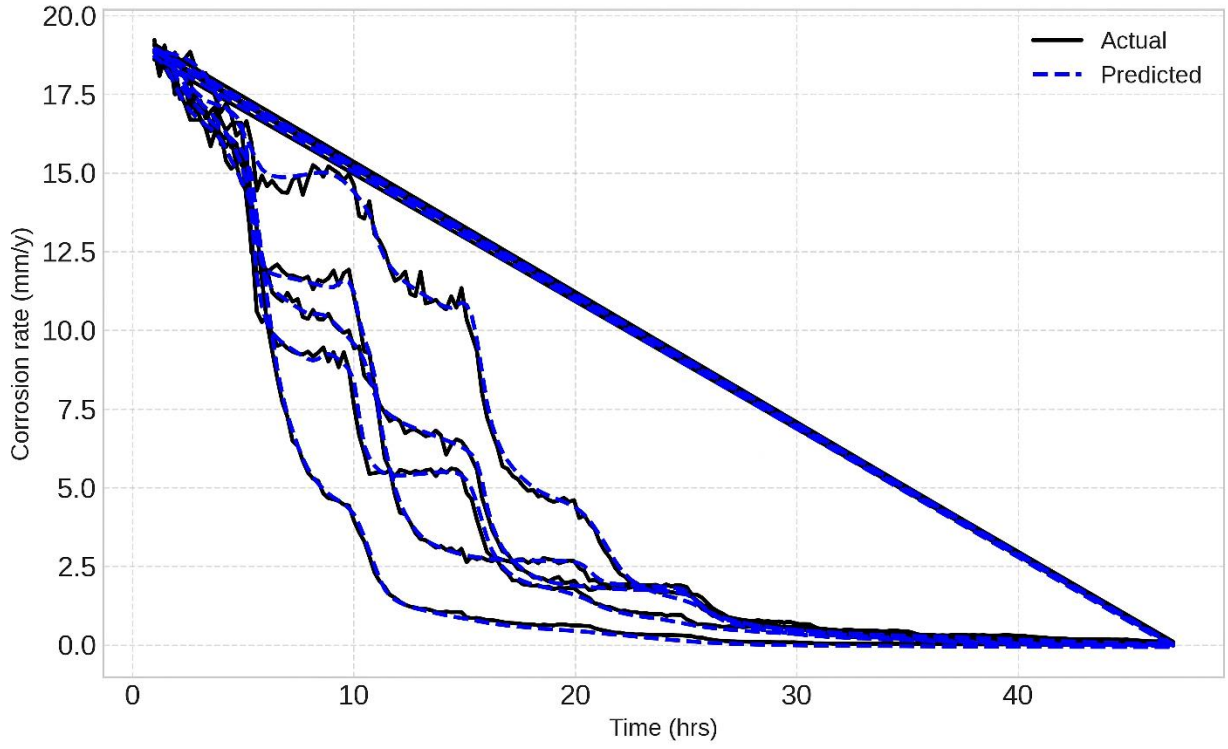

Figure 3. Time-series plot showing the predicted versus observed corrosion rate for the industrial case. The shaded band represents the 90% prediction interval (MC dropout)

Constraints: PGTT's physics prior depends on the mechanistic model's fidelity (miscalibration of the de Waard–Milliams model can bias results). In this process, seq_len = 24 is the best obtained result from this dataset.

10.57647/spre.2026.1001.01

Table 3. Ablation study (mean ± std across 5 run results)

Variant	MSE	R^2
PGTT (full)	0.067 ± 0.005	0.997 ± 0.001
w/o Physics Loss ($\lambda = 0$)	0.322 ± 0.010	0.9880.002

Table 4. Computational cost comparison (mean ± SD across five runs)

Model	Training time (GPU h)	Inference latency (ms/sample)
RF	0.2 ± 0.1	0.1 ± 0.05
SVR	0.5 ± 0.2	0.3 ± 0.1
MLP	1.0 ± 0.3	0.4 ± 0.1
Transformer	1.5 ± 0.4	0.5 ± 0.1
PGTT (proposed)	2.0 ± 0.5	0.6 ± 0.1

5.4. Comparison with Literature

Compared to previously reported data-driven or hybrid corrosion predictors [3, 5, 7, 22], PGTT provides competitive predictive performance while additionally assuring mechanistic consistency and feature-level interpretability through SHAP [15]. This balance is essential for industrial adoption, where explanation and physical plausibility are required for safety-critical decisions [2].

5.5. Broader Effect

PGTT supports safer, more sustainable operations by improving corrosion forecasting, enabling optimized inhibitor dosage, reducing maintenance costs, and reducing accident risk [5]. Integration with digital twin frameworks and Industry 4.0 pipelines enables proactive integrity management aligned with net-zero sustainability strategies [2].

5.6. Constraints

Although the PGTT demonstrates strong predictive performance on the subject dataset, the results are inherently limited to controlled laboratory-scale conditions. Because these experiments are designed to represent oil and gas pipeline environments, they cannot fully capture the complexity of real-world field operations, including multiphase flow behavior, extreme environmental conditions, and large-scale operational variability; consequently, further validation is required by applying long-term field data to assess the generalizability and robustness of this proposed framework in industrial-scale applications.

6. Conclusion

A PGTT with SHAP interpretability is proposed here. By running 22 inhibitor-dosing experiments, this PGTT yields MSE $= 0.067 \pm 0.005$, MAE $= 0.145 \pm 0.004$, and $R^2 = 0.997 \pm 0.001$ across five runs. The 90% predictive interval covers ~92% of the observed points. Removing the physics term degraded performance (MSE ≈ 0.32), revealing that the physics guidance is of great concern. PGTT is accurate and physically acceptable, making it appropriate for corrosion monitoring and inhibitor control in safety-critical infrastructure(s).

A PGTT, which integrates Transformer-based temporal modeling, a physics-informed loss derived from mechanistic corrosion understanding, and SHAP-based explainability, is proposed here. By applying this inhibitor-dosing dataset, PGTT yields MSE $= 0{,}067 \pm 0{,}005$, MAE $= 0{,}145 \pm 0{,}004$, and $R^2 = 0{,}997 \pm 0{,}001$ across five independent runs, while producing well-calibrated uncertainty estimates (90% prediction intervals entrapping approximately 92% of observations). Ablation results indicate that inclusion of the physics term substantially improves robustness ($w/o\ physics$: $MSE = 0{,}3218$). This PGTT provides a practical, interpretable, and physics-aware solution for CO_2 corrosion forecasting.

Data and Code Availability

The dataset applied in this study (SequentialDataInhibitor_large) and the source code of the Physics-Guided Temporal Transformer (PGTT) are openly available on GitHub at:
Source code and dataset:
https://github.com/Drafshin1357/-CO2-Induced-Corrosion-Rate-Predictions

All scripts for data preprocessing, model training, evaluation, and SHAP explainability are implemented in Python (PyTorch 2.0). Researchers can reproduce all results (MSE, MAE, R^2, and SHAP analyses) by running the provided Jupyter notebooks in the /notebooks directory. Additional experimental metadata is available upon request from the corresponding author.

Authors Contribution
All the authors have participated sufficiently in the intellectual content, conception and design of this work and the analysis and interpretation of the data, as well as the writing of the manuscript.

Availability of data and materials
The data that support the findings of this study and the source code are available at https://github.com/Drafshin1357/-CO2-Induced-Corrosion-Rate-Predictions.

Conflict of interests
The author states that there is no conflict of interest

References

[1] C. d. W. D. E. Milliams, "Carbonic acid corrosion of steel," *Corrosion*, vol. 31, no. 5, May 1975 1975, DOI:https://doi.org/10.5006/0010-9312-31.5.177

[2] M. M. H. J. Imran, S.; Mohamad Ayob, A. F, "A critical review of machine learning algorithms in maritime, offshore and oil & gas corrosion research: A comprehensive analysis of ANN and RF models," *Ocean Engineering*, vol. 295, March 2024 2024, DOI:https://doi.org/10.1016/j.oceaneng.2024.116796

[3] J. C. Fang, X.; Gai, H.; Lin, S.; Lou, H. H, "Development of machine learning algorithms for predicting internal corrosion of crude oil and natural gas pipelines," *Computers & Chemical Engineering*, vol. 177, p. 108 358, July 2023 2023. DOI:https://doi.org/10.1016/j.compchemeng.2023.108358

[4] L. B. Z. Coelho, Dawei; Van Ingelgem, Yves; Steckelmacher, Denis; Nowé, Ann; Terryn, Herman, "Reviewing machine learning of corrosion prediction in a data-oriented perspective," *npj Materials Degradation*, vol. 6, p. 8, 2022, DOI: https://doi.org/10.1038/s41529-022-00218-4

[5] Z. Z. Dong, Min; Li, Weirong; Wen, Fenggang; Dong, Guoqing; Zou, Lu; Zhang, Yongqiang, "Development of a Predictive Model for Carbon Dioxide Corrosion Rate and Severity Based on Machine Learning Algorithms," *Materials*, vol. 17, no. 16, p. 4046, 14 August 2024 2024, DOI: https://doi.org/10.3390/ma17164046

[6] A. S. Vaswani, N.; Parmar, N.; Uszkoreit, J.; Jones, Llion; *Gomez*, A. N.; Kaiser, Ł.; Polosukhin, I, "Attention Is All You Need," *Advances in Neural Information Processing Systems (NeurIPS 2017)*, vol. 30, pp. 5998–6008, December 2017 2017, DOI:https://doi.org/10.48550/arXiv.1706.03762

[7] J. W. Jiang, X.; Zhu, F.; Xiang, D.; Hu, Z.; Mu, S, "A deep learning framework integrating Transformer and LSTM *architectures* for pipeline corrosion rate forecasting," *Computers & Chemical Engineering*, vol. 190, p. Article No. 109365, 2025, DOI:https://doi.org/10.1016/j.compchemeng.2025.109365

[8] M. P. Raissi, P.; Karniadakis, G. E, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational Physics*, vol. 378, pp. 686–707, February 2019 2019, DOI: https://doi.org/10.1016/j.jcp.2018.10.045

[9] A. V. Dourado, Felipe A. C, "Physics-Informed Neural *Networks* for Corrosion-Fatigue Prognosis," *Proceedings of the Annual Conference of the PHM Society*, vol. 11, 2019, DOI: https://doi.org/10.36001/phmconf.2019.v11i1.814

[10] N. M. Chen, Rujin; Chen, Airong; Cui, Chuanjie, "PC-PINNs: Physics-informed neural networks for solving the phase-field model of pitting corrosion," *Bridge Maintenance, Safety, Management, Digitalization and Innovation*: CRC Press / Taylor & Francis, 2024, pp. 183–189. DOI:https://doi.org/10.1201/9781003483755-25

[11] X. L. Ma, Zhenzhi; Zhan, Jie; et al, "Prediction method of CO_2 plume distribution based on physics-informed neural networks," *Petroleum Drilling & Production Technology*, vol. 46, no. 5, pp. 90–98, 2024. DOI: https://doi.org/10.11911/syztjs.2024090

[12] A. W. enkatraman, Mark A.; Montes de Oca Zapiain, David, "Accelerating charge estimation in molecular *dynamics* simulations using physics-informed neural networks: corrosion applications," *npj Computational Materials*, vol. 11, no. 1, 2025, DOI: https://doi.org/10.1038/s41524-024-01495-0

[13] J. W. Jiang, X.; Zhu, F.; Xiang, D.; Hu, Z.; Mu, S, "A deep *learning* framework integrating Transformer and LSTM architectures for pipeline corrosion rate forecasting," *Computers & Chemical Engineering*, vol. 190, p. 109365, 2025, DOI:https://doi.org/10.1016/j.compchemeng.2025.109365

[14] S. M. L. Lundberg, S.-I., "A unified approach to *interpreting* model predictions," presented at the Advances in Neural Information Processing Systems (NeurIPS 2017), 2017. DOI: https://doi.org/10.48550/arXiv.1705.07874

[15] Y. S. Q. W. X. Z. L. D. S. B. D. Z. Z. Z. H. Z. Y. Xi, "*Interpretable* machine learning for maximum corrosion depth and influence factor analysis," *npj Materials Degradation*, vol. 7, no. 1, 2023. DOI: https://doi.org/10.1038/s41529-023-00324-x

[16] Q. H. Chen, W.; Wen, Y.; Wang, B.; Liu, G, "Degradation prediction and explainable analyses of the corroded subsea *pipelines* based on INSGAII-PGNN and SHAP algorithm," *Ocean Engineering*, vol. 334, p. Article no. 121495, 2025/08/01 (Aug 2025) 2025, DOI: https://doi.org/10.1016/j.oceaneng.2025.121495

[17] B. A. Lim, S. Ö.; Loeff, N.; Pfister, T, "Temporal Fusion Transformers for interpretable multi-horizon time series *forecasting*," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, October 2021 2021, DOI: https://doi.org/10.1016/j.ijforecast.2021.03.012

[18] L. Y. Tan, Y.; Zhang, K.; Liao, K.; He, G.; Tian, J.; Lu, X, "*Prediction* of internal corrosion rate for gas pipeline: A new method based on Transformer architecture," *Computers & Chemical Engineering*, vol. 198, p. Article No. 109084, 2025, DOI: https://doi.org/10.1016/j.compchemeng.2025.109084

[19] I. M. S. Alqahtani, A.; Khan, M, "Experimental and theoretical aspects of crack assisted failures of metallic *alloys* in corrosive environments – A review," *Materials Today: Proceedings*, vol. 66, pp. 2530–2535, July 2022 2022, DOI: https://doi.org/10.1016/j.matpr.2022.07.075

[20] M. M. Cid Montoya, A.; Verma, S.; Mures, O. A.; Rubio-*Medrano*, C. E, "Aeroelastic force prediction via temporal fusion transformers," *Computer-Aided Civil and Infrastructure Engineering*, vol. 39, pp. 319–344, January 2024 2024, DOI: https://doi.org/10.1111/mice.13381

[21] R. W. U. Revie, H. H, *Corrosion and Corrosion Control: An Introduction to Corrosion Science and Engineering (4th ed.)*. *John* Wiley & Sons, 2008. DOI: https://doi.org/10.1002/9780470277270

[22] G. T. He, J.; Sun, L.; Zhao, F.; Li, S.; Li, C.; Liao, K.; Chen, X.; Yang, W., "Research on natural gas pipeline corrosion *prediction* by integrating extreme gradient boosting and generative adversarial network," *SSRN Electronic Journal*, 2025, DOI: https://doi.org/10.2139/ssrn.5489826