



Machine learning-assisted intelligent interpretation of distributed fiber optic sensor data for automated monitoring of pipeline corrosion

Yiming Liu, Xiao Tan, Yi Bao^{*}

Department of Civil, Environmental and Ocean Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA

ARTICLE INFO

Keywords:

Automatic sensor data interpretation
Bayesian optimization
Distributed fiber optic sensor
Machine learning
Pipeline
Structural health monitoring

ABSTRACT

Distributed fiber optic sensor (DFOS) offers unique capabilities of monitoring corrosion for long pipelines. However, manually interpreting DFOS data is labor-intensive and time-consuming. To address this challenge, this paper presents a machine learning approach for real-time automatic interpretation of DFOS data used to monitor both uniform and non-uniform corrosion in pipeline. A machine learning model is developed to automatically detect corrosion based on DFOS data, and a corrosion quantification method is developed based on the output of the machine learning model. The proposed approaches are evaluated using laboratory experiments in terms of accuracy and robustness to pipeline diameter, spatial resolution of DFOS, type of fiber optic cable, and sensor installation methods. The results show that the F1 score for corrosion detection and the R^2 value for corrosion quantification are 0.986 and 0.953, respectively. This research will facilitate pipeline corrosion monitoring by enabling automatic distributed sensor data interpretation.

1. Introduction

Corrosion poses a significant challenge to the life-cycle performance of civil infrastructure as it causes substantial expenses in the maintenance of infrastructure [1]. According to a recent report by the National Association of Corrosion Engineers (NACE), the annual global cost of corrosion is estimated at a staggering \$2.5 trillion, accounting for 3.4 % of the global gross domestic product [2]. Beyond the considerable economic impact, corroded steel can jeopardize the integrity of reinforced concrete, prestressed concrete, and steel structures such as bridges and pipelines, potentially resulting in structural malfunction or even collapse. Based on data from the Pipeline and Hazardous Materials Safety Administration (PHMSA), corrosion is responsible for 11 % of natural gas pipeline incidents over the past three decades approximately [3]. There is an urgent need for real-time corrosion monitoring and timely assessment to mitigate these risks for the health of pipes and other structures [4].

Traditional methods for monitoring pipeline corrosion involve using different types of sensors. These existing methods often rely on electrochemical measurements such as potentiodynamic polarization [5], polarization resistance [6], electrochemical impedance [7], electrical resistance [8], and magnetic flux leakage [9]. These methods offer insights into the potential occurrence of corrosion at different cross

sections of a pipeline. These electrochemical sensing technologies remain the prevailing corrosion monitoring approach, directly gauging metal corrosion situations. However, due to their localized nature, these sensors and techniques require numerous installations or measurements along the pipeline, rendering the corrosion monitoring system complex and costly [10]. Furthermore, they are susceptible to electromagnetic interference and influenced by various environmental and pipeline-related variables [11]. The performance of these methods is compromised by alternating wet-dry conditions and seasonal weathering [12]. In addition to the electrochemical measurements, another approach involves using sound waves to detect corrosion [13]. This ultrasonic wave method sends high-frequency sound waves through the pipeline and analyzes the returning echoes. Although this method can provide a general idea of where corrosion might be occurring, the accuracy is limited [10].

Alternatively, distributed fiber optic sensors (DFOSSs) are garnering increasing attention for corrosion monitoring [14]. DFOSSs offer distinct advantages over traditional techniques, including immunity to electromagnetic interference, high precision, compact size, and robust physical and chemical stability [15]. Notably, a remarkable feature of DFOSSs is the utilization of a single fiber optic cable as both the transmission line and a sensor with numerous sensing points [16]. This unique feature eliminates the need for predicting corrosion locations, as DFOSSs

^{*} Corresponding author.

E-mail address: yi.bao@stevens.edu (Y. Bao).

<https://doi.org/10.1016/j.measurement.2024.114190>

Received 29 October 2023; Received in revised form 22 December 2023; Accepted 16 January 2024

Available online 19 January 2024

0263-2241/© 2024 Elsevier Ltd. All rights reserved.

continuously measure strains and temperature along their entire length. The above exceptional capabilities have led to diverse applications in the field of corrosion monitoring. For instance, corrosion-induced pipe wall thickness reduction was tracked through hoop strain measurement [17], and non-uniform corrosion in steel piles was monitored [18].

However, it is difficult to interpret the data measured from DFOSs deployed in large-scale pipelines. The following challenges concerning real-life applications have been identified: (1) The amount of data is so substantial in both spatial and temporal dimensions that automated monitoring is a challenge to achieve using traditional manual data processing and interpretation. In addition, skilled engineers are required to process DFOS data, which increases operation costs of DFOSs. (2) Existing methods of corrosion quantification assume pipe corrosion as a uniform corrosion along the circumference. Local corrosion situations require further consideration.

With the rapid development of machine learning techniques, several innovative methods have emerged for the analysis of strain distributions collected from DFOSs. For example, Song et al. [19,20] and Liu et al. [21,22] developed machine learning approaches to analyze DFOS data for monitoring cracks, automatically. The processing time for a substantial DFOS dataset containing 10,000 measurement points was less than 0.05 s [22]. These endeavors clearly highlight the promising potential of machine learning approaches for automated interpretation and efficient management of voluminous DFOS data. However, currently, there is no research exploring the automatic interpretation of DFOS data for corrosion monitoring using machine learning.

To address the above challenge, this paper presents an approach for automatic interpretation of DFOS data for monitoring pipeline corrosion. This paper has three main contributions: (1) A machine learning model is developed by integrating a convolutional neural network (CNN), a long short-term memory (LSTM) module, an attention module, and a fully connection module, aiming to detect pipeline corrosion from DFOS data. The dimensions of input data and hyperparameters of the

machine learning model are optimized via Bayesian optimization. (2) Based on the machine learning model, a method is proposed to quantify both uniform and non-uniform corrosion. (3) The performance of the proposed approach and important influencing factors are evaluated based on laboratory corrosion experiments. The investigated factors include the pipeline diameter, spatial resolution of DFOS, type of fiber optic cable, and sensor installation methods.

The remainder of the paper is organized as follows: Section 2 introduces the methods. Section 3 introduces the laboratory experiments. Section 4 presents and discusses the experimental results. Section 5 summarizes the conclusions.

2. Methods

2.1. Overview

The framework of the proposed approach is shown in Fig. 1, consisting of four main steps: (1) A DFOS and an optic analyze system are used to collect strain values on the surface of pipes (Section 2.2). (2) The collected strain values are input into a machine learning model to determine whether corrosion occurs. (3) A Bayesian optimization-based hyperparameter tuning method is developed to enhance the performance of the proposed machine learning model. (4) Based on the results of corrosion detection from the machine learning model, a method is proposed to quantify corrosion in uniform and non-uniform corrosion cases.

2.2. Distributed fiber optic sensors

In this study, a single-mode fiber optic cable (model: Corning® SMF-28e+® [23]) was used as the distributed sensor and transmission line. The cable consisted of a fused silica fiber core, a fused silica fiber cladding, and multiple layers of polymer coatings for mechanical

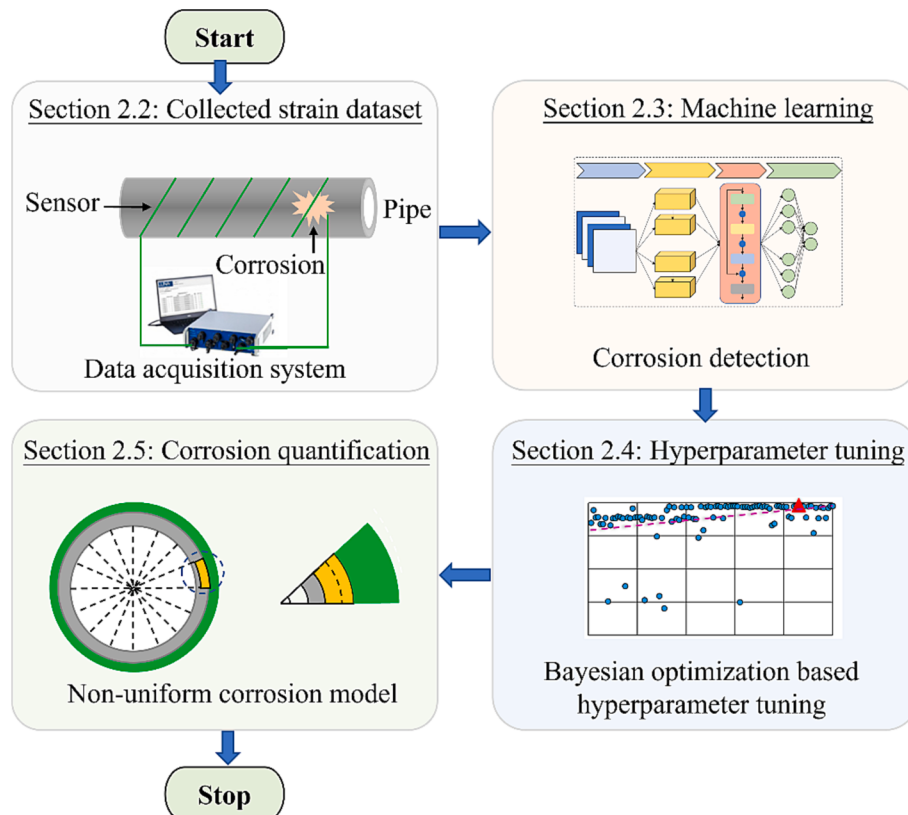


Fig. 1. Framework of the proposed approach for interpreting DFOS data via machine learning.

protection. Light waves propagate in the fiber core via total internal reflection at the core-cladding interface [24]. The fiber optic cable is attached to the surface of a pipeline. When the pipeline is subjected to mechanical loads or corrosion that generates strain changes, the propagation of light in the fiber will change accordingly. The strains in the fiber optic cable is measured with an optical frequency domain reflectometry (OFDR) technique based on a calibration relationship between strain and frequency shift [10].

In this study, a Luna ODiSI 6120 was used to measure strains based on the OFDR technique [25]. A DFOS was attached to the surface of a pipe in a helix pattern for monitoring the corrosion of the pipe, as shown in Fig. 2(a). The radius of the pipe is R_0 . The number of loops of fiber optic cable on the surface of the pipe is m . Based on the OFDR technique, the collected strain values at the time of t can be stored as a vector:

$$\boldsymbol{\epsilon}^t = [\epsilon_{1,1}^t \ \epsilon_{1,2}^t \ \dots \ \epsilon_{1,j}^t \ \dots \ \epsilon_{1,n}^t \ \epsilon_{2,1}^t \ \epsilon_{2,2}^t \ \dots \ \epsilon_{2,n}^t \ \dots \ \epsilon_{i,j}^t \ \dots \ \epsilon_{m,n}^t] \quad (2)$$

where ϵ_{ij}^t denotes the j -th strain value in the i -th loop of the DFOS at time t . The total number of strain data in each loop is n , which can be calculated by:

$$n = \left\lfloor \frac{\sqrt{(2\pi R_0)^2 + h^2}}{S} \right\rfloor \quad (3)$$

where h is the winding spacing of the DFOS on the surface of the pipe, S is the spatial resolution of the DFOS, and $\lfloor \cdot \rfloor$ denotes a round down function.

Within a period of time t , the DFOS collects $m \times n \times t$ strain data, which can be used to monitor pipe corrosion. As time progresses, the

volume of DFOS data increases rapidly. For example, assuming $R_0 = 20$ mm, $h = 20$ mm, $S = 0.65$ mm, $m = 1000$, and $t = 100$, the DFOS will collect a total of $1000 \times \left\lfloor \frac{\sqrt{(2\pi \cdot 20)^2 + 20^2}}{0.65} \right\rfloor \times 100 = 19,500,000$ data. The data volume poses a significant challenge for manual processing and interpretation. Numerous data are collected, and the data change in both spatial and temporal dimensions, making it difficult to distinguish red dots from blue dots, as shown in Fig. 2(b).

2.3. Machine learning model

A machine learning architecture integrating a CNN, an LSTM module, a self-attention module, and a fully connection module is proposed in this study, as shown in Fig. 3. Previous research has shown that CNNs can efficiently extract spatial features from spatial data [26], and LSTM can effectively extract temporal features from sequential data [27]. This architecture has components to extract both spatial and temporal features from the DFOS data that are collected over time. The input and output data of the machine learning model is presented in Section 2.3.1. The modules are elaborated in the following subsections.

2.3.1. Input and output data

In this study, strain distributions along DFOS were measured based on the OFDR technique. The details of data collection are presented in Section 3. The strain distributions were segmented in spatial and temporal dimensions using a sliding window method [28], as shown in Fig. 4.

The window sizes in the spatial and temporal dimensions are noted as A and B , respectively. As the window slides along the spatial and

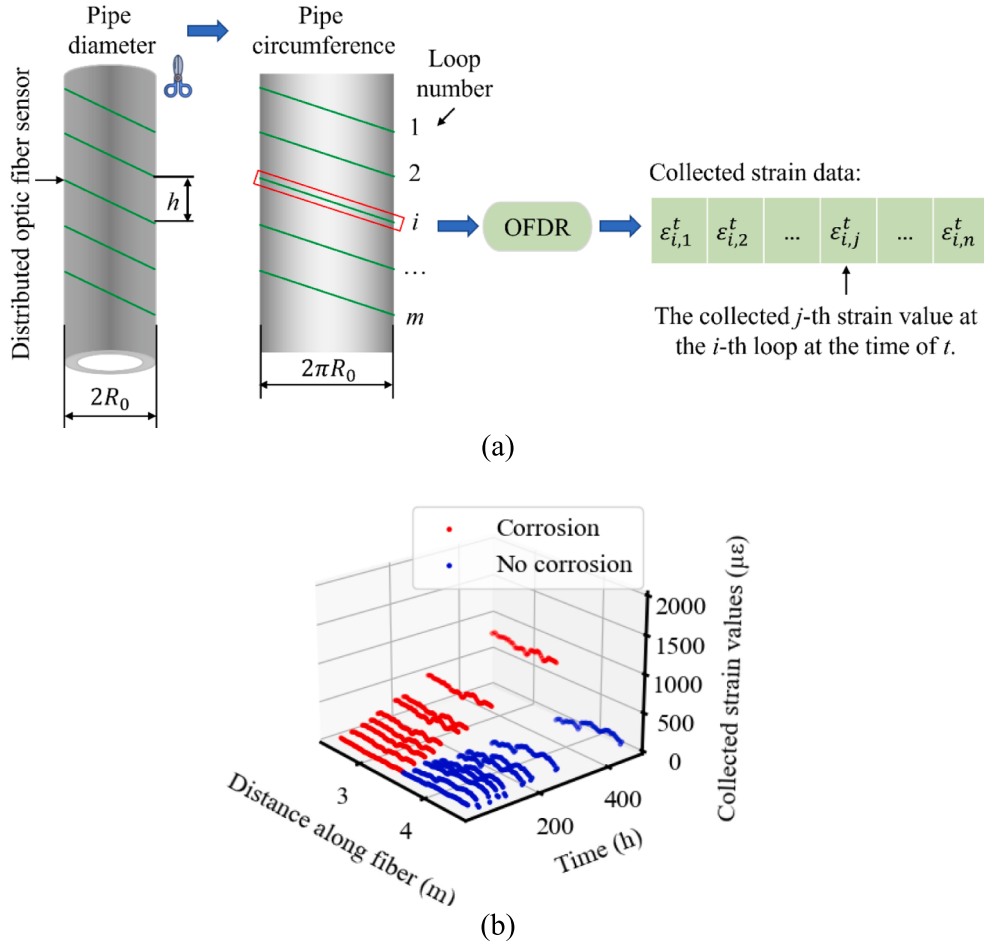


Fig. 2. Strain data acquisition for monitoring pipeline corrosion based on DFOS and OFDR: (a) illustration of the process; and (b) example of real collected data.

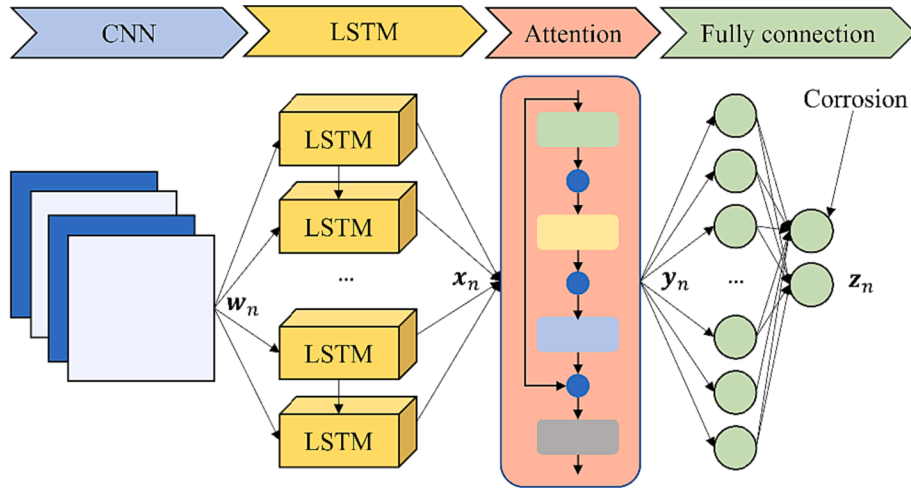


Fig. 3. Architecture of the proposed machine learning model for analyzing DFOS data.

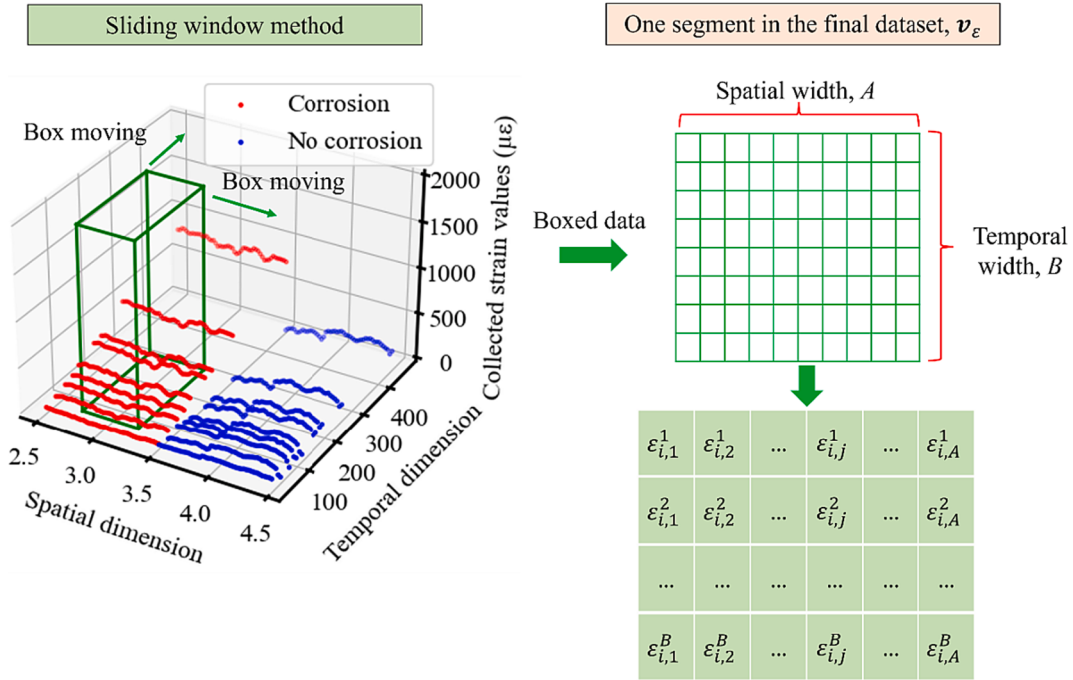


Fig. 4. Generation of data segments using a sliding window in spatial and temporal dimensions.

temporal dimensions, a set of data is segmented and stored in a box that has A units in length and B units in width. The height of the box is the strain value. When the data corresponds corrosion, the data is labeled as 1. When the data corresponds no corrosion, the data is labeled as 0. Each data segment is a two-dimensional matrix, denoted as v_e . These segments form the input data for the proposed machine learning method. The optimal window sizes are discussed in Section 4.1. The output data z_e of this machine learning method is corrosion condition.

2.3.2. Convolutional neural network

A CNN is employed to extract spatial features from the data segment, v_e . The CNN employs K one-dimensional spatial convolution layers to capture a spatial feature that can effectively represent the spatial strain values collected from the DFOSs. Each convolutional layer takes the data segment v_e as input, yielding an output defined as:

$$w_{e,k} = v_e \otimes a_k + b_k \quad (4)$$

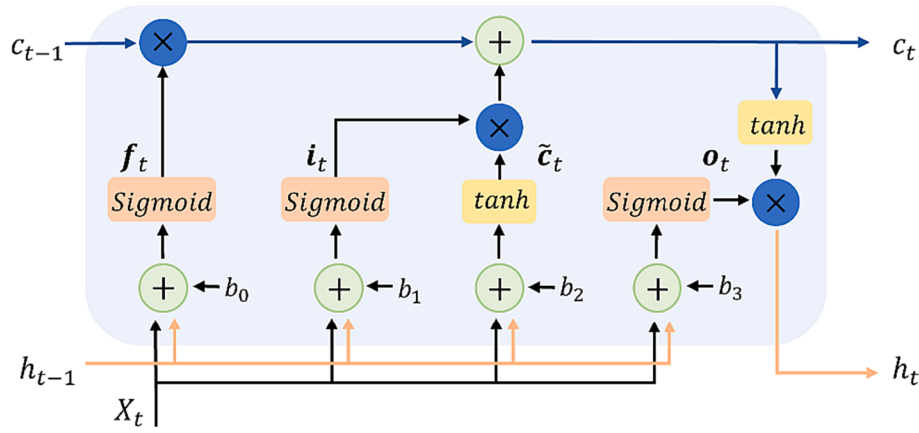
where $w_{e,k}$, a_k , and b_k are the output, weight, and bias of the k -th convolutional layer, respectively, and $\otimes$ represents the convolution operation. The resulting dimension of the output w_e from the complete CNN module is $B \times K$.

2.3.3. Long short-term memory module

Designed with B standard LSTM cells and H ($H \geq 1$) hidden units, an LSTM module is developed to extract temporal features from the output of the previous CNN module. A standard LSTM cell at time step of t is illustrated in Fig. 5, where the functions $\text{sigmoid}(\bullet)$ and $\text{tanh}(\bullet)$ are the sigmoid and hyperbolic tangent functions, respectively. Equations (5) to (10) present the mathematical formulation of the standard LSTM cell [29,30].

$$f_t = \text{sigmoid}(W_{fx} \times X_t + W_{fh} \times h_{t-1} + b_0) \quad (5)$$

$$i_t = \text{sigmoid}(W_{ix} \times X_t + W_{ih} \times h_{t-1} + b_1) \quad (6)$$

Fig. 5. Structure of a standard LSTM cell of the machine learning model at time t .

$$\tilde{c}_t = \tanh(W_{cx} \times X_t + W_{ch} \times h_{t-1} + b_2) \quad (7)$$

$$o_t = \text{sigmoid}(W_{ox} \times X_t + W_{oh} \times h_{t-1} + b_3) \quad (8)$$

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \quad (9)$$

$$h_t = o_t \times \tanh(c_t) \quad (10)$$

where f_t , i_t , and o_t correspond to the forget gate, input gate, and output gate of the standard LSTM cell at time step of t , respectively. X_t , c_t , and h_t represent the current input, memory cell, and hidden state, respectively. W and b with distinct subscripts denote the corresponding weights of different gates. The dimension of the output x_e from the entire LSTM module is $B \times H$, which is determined by the number of LSTM cells and the length of the hidden state.

2.3.4. Attention module

Attention mechanism is a fundamental concept in natural language processing and computer vision, which enables machine learning models to focus on specific parts of input data while processing it. Alongside the well-known scaled dot-product attention utilized in transformer models [31], there exists a range of other prominent attention types, such as content-based attention [32], general attention [33], and location-based attention [33].

A local attention mechanism is utilized (Fig. 6). The input data for the attention module is derived from the output of the LSTM module, and the output (y_e) is flattened to generate a one-dimensional vector with dimensions of $1 \times (H \times S)$:

$$y_e = \text{sigmoid}(W_2 \times \tanh(W_1 \times x_e^T)) \times x_e \quad (11)$$

where W_1 and W_2 are the trainable weight matrices with two parameters of attention length L and output size S ; and x_e^T is the transposed input data.

2.3.5. Fully connection module

A fully connected module is often referred to a dense layer in neural networks and establishes comprehensive connections between input and output nodes. In this module, the input vector y_e undergoes a linear transformation by employing a weight matrix W_z and a bias term b_4 .

Table 1

Candidates of the undetermined hyperparameters.

Hyperparameter	Number of candidates	Candidates
Number of CNN filters	63	{2, 3, ..., 64}
Number of LSTM units	63	{2, 3, ..., 64}
Number of hidden units in the attention	15	{2, 3, ..., 16}
Number of distinct components in the attention	15	{2, 3, ..., 16}
Learning rate	>100	$[10^{-5}, 10^{-3}]$ (continuous)
Batch size	6	{2, 4, 8, 16, 32, 64}

Subsequently, a non-linear activation function $\text{sigmoid}(\bullet)$ is applied to generate the final result of the entire machine learning model z_e . The output z_e is one-dimensional vector with a dimension of 1×2 , indicating the corrosion detection results.

$$z_e = \text{sigmoid}(W_z \times y_e + b_4) \quad (12)$$

2.4. Hyperparameter tuning

The proposed machine learning architecture has hyperparameters that have a significant effect on the performance of machine learning model. These hyperparameters are shown in Table 1. Regarding the architecture, the hyperparameters include the number of CNN filters, the number of LSTM units, the number of hidden units to decode the attention, and the number of distinct components in the attention module. Regarding the learning process, the hyperparameters include the learning rate and batch size. When hyperparameter tuning is performed using the grid search method, the method will require iterations for $63 \times 63 \times 15 \times 15 \times 100 \times 6 = 535,815,000$ times. Alternatively, random search can be employed, but it introduces significant variance due to its reliance on randomly selected parameters. In addition to the above limitations, both the grid search and random search methods lack the ability to learn from prior search results. To overcome these limitations, a Bayesian optimization-based hyperparameter tuning technique is proposed. This approach leverages Bayesian principles to transform previously explored parameters into insightful prior

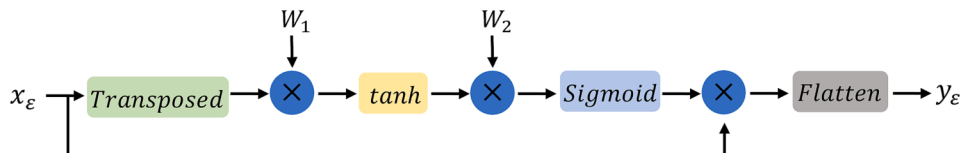


Fig. 6. Structure of the attention module incorporated into the machine learning architecture.

information.

The Bayesian optimization-based hyperparameter tuning can be expressed as follows:

$$\theta^* = \underset{\theta \in \Theta}{\operatorname{argmin}} \Psi(\theta) \quad (13)$$

where $\Psi(\theta)$ is the objective function (or loss function) of the proposed machine learning model; θ denote hyperparameter values; Θ is the search space (or candidates) of the hyperparameters, as shown in Table 1; and θ^* is the best hyperparameter that minimizes the output of $\Psi(\theta)$.

Because each computation of the objective function $\Psi(\theta)$ in Equation (13) is a training of the machine learning model, its time consumption is often excessive. To mitigate this computational challenge, Bayesian optimization-based hyperparameter optimization techniques are leveraged. These methods employ surrogate models and acquisition functions to streamline the process. A Gaussian process is used to establish a surrogate model for the objective function:

$$\Psi(\theta) \sim \mathcal{GP}(\mu(\theta), k(\theta, \theta')) \quad (14)$$

where $\mathcal{GP}(\bullet)$ denotes the Gaussian process, $\mu(\bullet)$ represents the mean function associated with different hyperparameter combinations, and $k(\bullet)$ is the covariance function between two hyperparameter combinations.

The Gaussian process integrates information from the current hyperparameter combination and combinations, i.e., $p(\Psi(\theta_{t+1})|\theta_{t+1}, \Psi_t)$, where t is the t -th hyperparameter combination. For each input hyperparameter combination, the surrogate model outputs a Gaussian distribution rather than the value of the objective function. In general, for hyperparameter tuning of machine learning models, the mean of the Gaussian distribution is considered as $\mu(\theta) = 0$, and the covariance function is computed using the Matern kernel function. The objective of the acquisition function is to maximize the probability that the next hyperparameter combination achieves a higher objective function. This is carried out to maximize the probability of improvement (POI):

$$POI(\theta_{t+1}) = P(\Psi(\theta_{t+1}) \geq \Psi(\theta^+) + \xi) = \Phi\left(\frac{\mu(\theta_{t+1}) - \Psi(\theta^+) - \xi}{\sigma(\theta_{t+1})}\right) \quad (15)$$

where θ^+ denotes the current combination of hyperparameter values that can minimize the objective function; $\Phi(\bullet)$ is the normal cumulative distribution function; ξ is a trade-off parameter greater than or equal to 0; $\mu(\theta_{t+1})$ and $\sigma(\theta_{t+1})$ denote the corresponding mean and standard deviation of the hyperparameter combination is θ_{t+1} , respectively.

As ξ approaches 0, the next hyperparameter combination θ_{t+1} approaches θ^+ , meaning that this hyperparameter tuning tends to exploit the most-likely region of the global optimum according to the known posterior distributions. On the contrary, this hyperparameter tuning explores the unknown region. Through continuous iterations, when the stopping condition is reached, the current hyperparameter combination θ^+ is the combination of hyperparameter values θ^* that can minimize the output of the objective function.

2.5. Quantification of corrosion

In the course of steel corrosion, as more and more rust is produced and accumulates on the surface of pipeline, the overall diameter of the pipeline is increased because rust is porous and has a larger volume than the original steel. The increase of the pipeline diameter exerts tensile strains to the fiber optic cable attached to the exterior surface of the pipeline, as shown in Fig. 7. The original diameter of the pipeline is R_0 . Due to corrosion, the thickness of steel is reduced to R_w . The radius of the pipe with rust is noted as R_c . The radius of the pipe with rust and the thickness of fiber optic cable is R_f .

Curves L_1 and L_2 indicate the edges of pipeline before and after

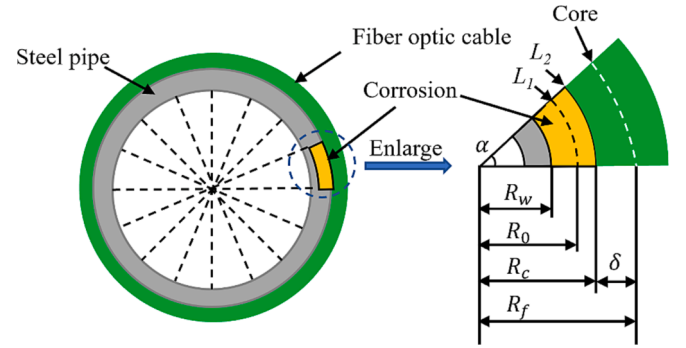


Fig. 7. Non-uniform corrosion model of a steel pipe with a fiber optic cable on the surface.

corrosion, respectively. According to Fig. 7, the following equation is satisfied:

$$\alpha = \frac{L_1}{R_0} = \frac{L_2}{R_c} \quad (16)$$

The volumes of consumed steel (V_p) and corrosion rust (V_r) between two loops of the DFOS can be expressed as:

$$V_p = \frac{\alpha}{2\pi} \pi h (R_0^2 - R_w^2) = \frac{ah}{2} (R_0^2 - R_w^2) \quad (17)$$

$$V_r = \frac{\alpha}{2\pi} \pi h (R_c^2 - R_w^2) = \frac{ah}{2} (R_c^2 - R_w^2) \quad (18)$$

The j -th mass loss (Δm_{ij}^t) caused by corrosion in the i -th loop of the DFOS at time t can be expressed as:

$$\Delta m_{ij}^t = V_p \rho = \frac{ah\rho}{2} (R_0^2 - R_w^2) \quad (19)$$

where ρ is the density of pipe material ($\rho = 7850 \text{ kg/m}^3$ for steel).

The strain variation (ϵ_{ij}^t) due to the increase of pipeline diameter can be calculated by the average elongation from curve L_1 to curve L_2 . Due to the thickness (δ) of the DFOS, the measured strain values from the DFOS reflect the elongation of the core of the DFOS instead of the surface of the pipeline, although the DFOS is tightly attached to the surface of the pipeline. In practice, the thickness of the DFOS is negligible compared with the radius of the pipeline:

$$R_0 + \delta \approx R_0 \quad (20)$$

Further discussion on the effect of δ is presented in Section 4.2.3 and Section 4.3.2. The strain variation and R_c can be calculated as follows:

$$\epsilon_{ij}^t = \frac{\alpha(R_0 + \delta) - \alpha(R_c + \delta)}{\alpha(R_0 + \delta)} = \frac{R_0 - R_c}{R_0 + \delta} \approx \frac{R_0 - R_c}{R_0} \quad (21)$$

$$R_c = R_0(1 - \epsilon_{ij}^t) - \epsilon_{ij}^t \delta \approx R_0(1 - \epsilon_{ij}^t) \quad (22)$$

The volume expansion coefficient (λ) due to the expansion of corrosion rust is defined as:

$$\lambda = \frac{V_r}{V_p} = \frac{R_c^2 - R_w^2}{R_0^2 - R_w^2} \quad (23)$$

Substituting Equations (16), (21), (22), and (23) into Equation (19), the mass loss can be expressed as:

$$\Delta m_{ij}^t = \frac{L_1 h \rho R_0}{2(\lambda - 1)} \left[2\epsilon_{ij}^t + (\epsilon_{ij}^t)^2 \right] \quad (24)$$

When there are m loops of fiber optic cable and n points of measurement in each loop, the total mass loss Δm is expressed as:

$$\Delta m = \sum_{i=1}^m \sum_{j=1}^n \Delta m_{ij}^t z_{ij}^t = \sum_{i=1}^m \sum_{j=1}^n z_{ij}^t \frac{L_1 h \rho R_0}{2(\lambda - 1)} \left[2\epsilon_{ij}^t + (\epsilon_{ij}^t)^2 \right] \quad (25)$$

$$z'_{ij} = \begin{cases} 1, e'_{ij} \text{ represents corrosion} \\ 0, \text{otherwise} \end{cases} \quad (26)$$

where z'_{ij} is the corrosion result obtained from the machine learning model corresponding to e'_{ij} .

The main idea of the proposed corrosion quantification method is to divide the cross section of a pipeline into multiple parts. The proposed machine learning method is used to identify whether the strain distribution corresponding to each part represents corrosion or not. When corrosion is detected, the corresponding mass loss is calculated. The cumulative value of all mass losses is then the overall mass loss of the pipe. This method can tackle both local and global corrosion. Compared with previous research, which assumes uniform pipeline corrosion along the circumference [10], the proposed method can address non-uniform corrosion.

2.6. Performance metrics

Various performance metrics were proposed to evaluate machine learning models. Among them, F1 score has been commonly adopted [34]. F1 score combines precision and recall into a single value from 0 to 1. A perfect classifier with high precision and recall achieves a score of 1. The F1 score is well-suited for binary classification tasks [34].

For corrosion quantification, mass loss is usually employed to evaluate corrosion severity of pipelines [10,18]. The quantification performance is evaluated with the coefficient of determination (R^2) between the true values and calculated values of mass loss [27]. The R^2 value varies between 0 and 1, and 1 indicates that the corrosion quantification model perfectly evaluates the mass loss. The F1 score and R^2 are defined as:

$$F1 = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \quad (27)$$

$$R^2 = 1 - \frac{\sum_{k=1}^K (\Delta m_k - \widehat{\Delta m}_k)^2}{\sum_{k=1}^K (\Delta m_k - \bar{\Delta m})^2} \quad (28)$$

where TP indicating the number of the segments representing corrosion accurately identified as corrosion; FP denoting the number of the segments wrongly identified as corrosion; FN denoting the number of the segments representing corrosion wrongly identified as no corrosion; Δm_k and $\widehat{\Delta m}_k$ represents the true and calculated mass loss of the k -th pipe, respectively; K is the number of pipes; $\bar{\Delta m}$ denotes the average value of the true mass loss of the K pipes.

3. Experimental program

3.1. Experiments

In this research, three types of steel pipes were investigated. The pipes were immersed in a sodium chloride solution (concentration: 3.5 % by mass) for corrosion tests at room temperature (25 ± 2 °C). The

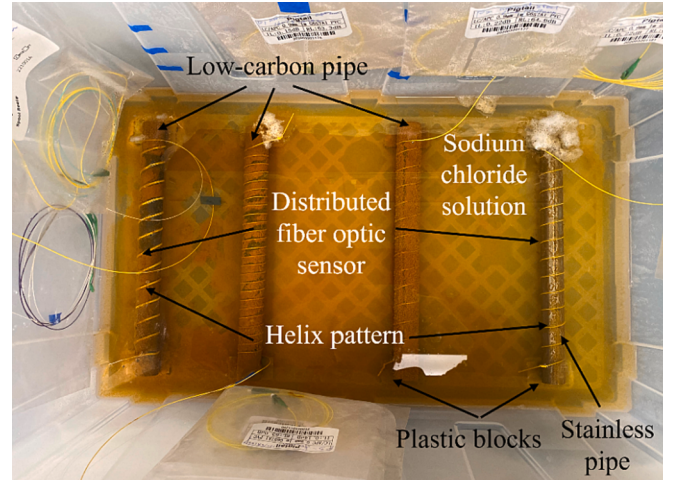


Fig. 9. Steel pipe specimens immersed in sodium chloride solution for corrosion experiments.

dimensions and materials of the pipes are shown in Fig. 8, including stainless pipes, small low-carbon steel pipes, and large low-carbon steel pipes.

DFOSs were attached to the exterior surfaces of the pipe specimens following a helix pattern, as shown in Fig. 9. The total length of the DFOS was around 4.4 m. Different spacings between adjacent rounds of spiral were investigated to optimize the sensor installation method. The pipes were supported by plastic blocks at the two ends to expose the pipes to the sodium chloride solution. The fiber optic cable was connected to the Luna ODISI system for measuring strain distributions.

The measurements were carried out at 12 different immersion times (0 h, 30 h, 60 h, 114 h, 134 h, 184 h, 208 h, 280 h, 472 h, 640 h, 912 h, and 1080 h). The gauge length of the DFOS was 0.65 mm [25]. Another distributed sensor was installed closely for temperature compensation.

The mass loss of the pipe specimens due to steel corrosion was measured to quantify corrosion. The initial mass of the pipe specimens was gauged with a high precision balance before immersing. At 12-hour intervals, a pipe specimen was extracted, rust was carefully eliminated using vinegar acid, and the mass after corrosion was measured using the high-precision balance. The difference between this mass and the corresponding initial mass represents the mass loss due to corrosion, which serves as the true mass loss to evaluate the proposed corrosion quantification method in Section 4.3. More details about the experiments can be found in Ref. [10].

3.2. Datasets

Based on the gauge length (0.65 mm), the 4.4-m DFOS measured $4400/0.65 = 6769$ strain data in each measurement and a total of $6769 \times 12 = 81,228$ strain data in 12 measurements. Since the stainless-steel pipes did not corrode, the strain distributions measured from the

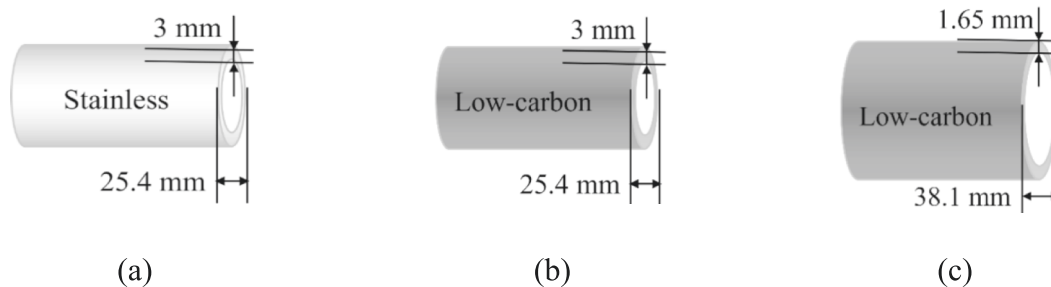


Fig. 8. Pipe specimens: (a) stainless pipes, (b) small low-carbon steel pipes, and (c) large low-carbon steel pipes.

stainless-steel pipes were used as the reference (no corrosion) data for evaluating the corrosion severity of the corroded specimens. The collected strain data representing no corrosion are shown in Fig. 10(a). The minor increase of tensile strains in the DFOS can be attributed to the polymer coatings of the fiber optic cable because the coatings can absorb water and swell. The strain distributions collected from the low-carbon steel pipe specimens are shown in Fig. 10(b). With the increase of immersion time, the strains of low-carbon steel pipes have shown a significant increase. The increase of strains can be attributed to the growth of rust on the surface of the pipe specimens.

In the laboratory corrosion experiment, the strain distributions were measured on a daily basis. In this study, linear interpolation was used to generate more data. Specifically, strain distributions were generated every 3 h via linear interpolation [35]. With the generated strain dataset, a total of 12 datasets were generated in both spatial and temporal dimensions using the sliding window approach (Section 2.3.1), as shown in Table 2.

Regarding the spatial dimension, the sliding width is determined by the fractions (1/16, 2/16, 3/16, and 4/16.) of the circumference of the pipe specimens. Specifically, the designated candidate lengths were 5 mm, 10 mm, 15 mm, and 20 mm, respectively. Regarding the time dimension, the sliding window widths include 3 h, 6 h, and 9 h. The effect and the optimization of the sliding width are elaborated in Section 4.1. Each dataset was split into a training dataset, a validation dataset, and a testing dataset with a ratio of 7:1:2. The data in the testing dataset was not used in the training of the machine learning model, and was used to verify the generalization ability of the model, i.e., the ability of the model to predict unknown data.

4. Results and discussion

4.1. Machine learning model

To determine the structure and hyperparameters of the machine learning model, the effects of hyperparameters and sliding widths on the accuracy of the model were investigated. The datasets in Table 2 were utilized to train the machine learning model and tune the hyperparameters based on Bayesian optimization. The results are summarized in Table 3. When the input data format is (23, 3), the F1 score of the machine learning model is 0.835 before performing hyperparameter tuning and 0.986 after performing hyperparameter tuning based on Bayesian optimization. The improvement indicates that hyperparameter tuning is important. With the increase of the number of optimization

Table 2

Summary of the dataset used in this study.

No.	Spatial length (mm)	Ratio between the spatial length and the circumference	Time length (hour)	Number of strain values		Dataset size
				Spatial	Temporal	
1	5	1/16	3	7	2	(7, 2)
2	5	1/16	6	7	3	(7, 3)
3	5	1/16	9	7	4	(7, 4)
4	10	2/16	3	15	2	(15, 2)
5	10	2/16	6	15	3	(15, 3)
6	10	2/16	9	15	4	(15, 4)
7	15	3/16	3	23	2	(23, 2)
8	15	3/16	6	23	3	(23, 3)
9	15	3/16	9	23	4	(23, 4)
10	20	4/16	3	30	2	(30, 2)
11	20	4/16	6	30	3	(30, 3)
12	20	4/16	9	30	4	(30, 4)

steps, the accuracy of the machine learning model increases overall, as shown in Fig. 11. The highest F1 score value occurs in the 86th step and is marked by a red triangle. The optimization process shows that the hyperparameter tuning based on Bayesian optimization can significantly improve the accuracy of the machine learning model.

The effects of the sliding widths (i.e., A and B in Section 2.3.1) on the accuracy of the machine learning model are shown in Fig. 12. With the increase of the sliding width in spatial dimension, as shown in Fig. 12(a), the accuracy of the machine learning model increases and then stabilizes, meaning that data from a minimum length of DFOS are necessary for extracting sufficient spatial features of the data to identify corrosion. It is recommended to set the minimum spatial dimension (i.e., A) at 15 mm. With the increase of the sliding width in temporal dimension, as shown in Fig. 12(b), the accuracy of the machine learning model

Table 3

Effects of the hyperparameters on the performance of the proposed model.

Hyperparameter	Before tuning	After tuning
Number of CNN filters	4	10
Number of LSTM units	8	56
Number of hidden units in the attention	8	2
Number of distinct components in the attention	8	12
Learning rate	0.0001	0.0001
Batch size	8	2
F1 score of the model	0.835	0.986

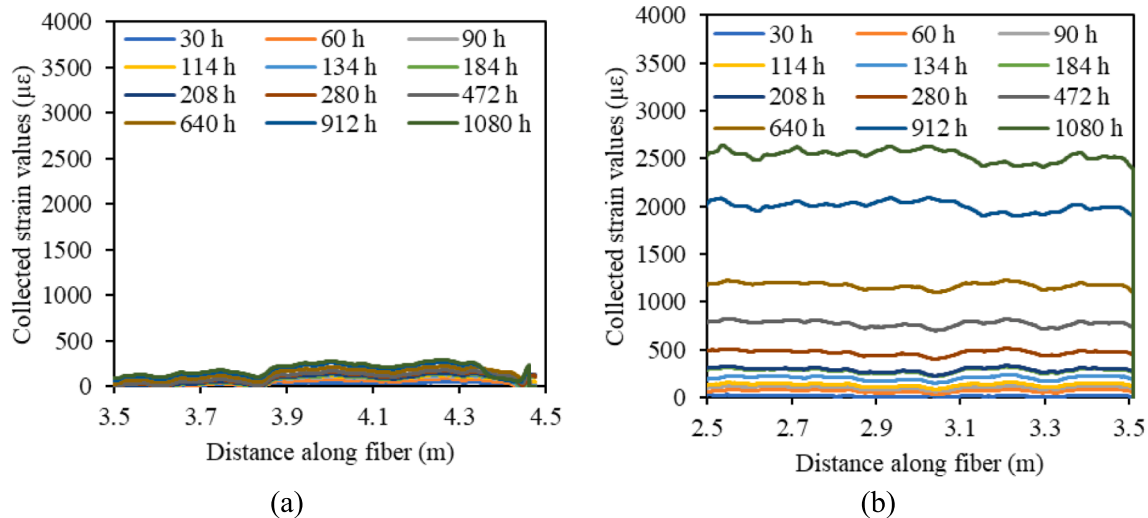


Fig. 10. Strain distributions measured from DFOSs on the surface of pipe specimens: (a) stainless-steel pipes (no corrosion) and (b) low-carbon steel pipes (corrosion).

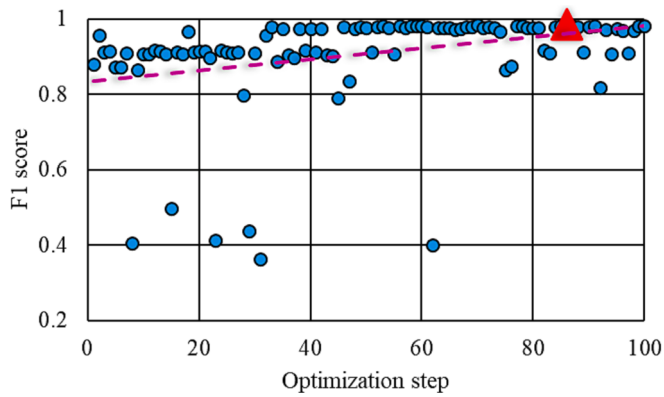


Fig. 11. Improvement of F1 score in the hyperparameter tuning based on Bayesian optimization.

increases and then stabilizes, meaning that data from a minimum period of measurements from DFOS are necessary for extracting sufficient temporal features of the data to identify corrosion. It is recommended to set the minimum temporal dimension (i.e., B) at 6 h.

The interactions between spatial and temporal dimensions are shown in Fig. 12. When the target F1 score is 0.95, the required spatial dimension and temporal dimension are dependent on each other. When the temporal dimension is 9 h, the F1 score exceeds 0.95, even when the spatial dimension is only 5 mm, as shown in Fig. 12(a). When the spatial dimension is 15 mm, the F1 score exceeds 0.95 as long as the temporal dimension is not shorter than 6 h, as shown in Fig. 12(b). In short, it is important to optimize the selection of spatial and temporal dimensions in specific applications. Thereafter, the spatial and temporal dimensions are 15 mm and 6 h, respectively.

4.2. Detection results and robustness

4.2.1. Detection results

The performance of the machine learning model is evaluated by its ability to identify segments representing corrosion and segments representing no corrosion accurately. This process involves four distinct scenarios (Fig. 13): (1) true positives (TP), indicating accurate identification of the segments representing corrosion; (2) true negatives (TN), where the model correctly identifies the segments representing no corrosion; (3) false negatives (FN), where actual corrosion is overlooked by the model; and (4) false positives (FP), denoting instances wrongly identified as corrosion.

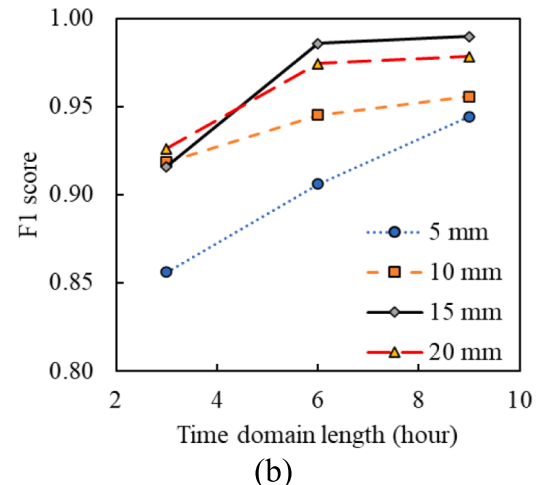
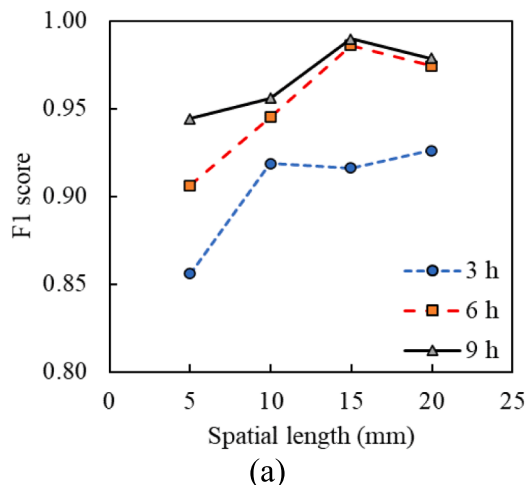


Fig. 12. Effect of (a) the spatial length and (b) the temporal length in the datasets on the accuracy of the trained model.

The confusion matrix in Fig. 14 provides a comprehensive overview of the performance of the proposed machine learning model. As shown in Fig. 14(a), the confusion matrix presents the counts of TP, TN, FN, and FP [34]. Each of these cases corresponds to a specific cell in the confusion matrix. The specific statistical results of the proposed model are shown in Fig. 14(b). This visual representation shows how well the model accurately predicts different cases. Notably, the F1 score reaches an impressive value of 0.986, further affirming the effectiveness of the proposed machine learning method in corrosion detection tasks.

4.2.2. Comparison with other machine learning models

The performance of the proposed machine learning model has been compared with that of other popular machine learning models, including support vector machine (SVM), extreme gradient boosting decision tree (XGBoost), multi-layer perceptron (MLP), LSTM, CNN, combined CNN and LSTM, and deep belief network (DBN). The investigated performance metrics include the accuracy and efficiency of the different machine learning models. The accuracy and efficiency are evaluated using the F1 score and average detection time, respectively, as shown in Fig. 15. The proposed model successfully identifies corrosion and no corrosion from a total of 2000 strain distributions with an F1 score of 0.986 within 1.2 s, outperforming the other investigated machine learning models. Traditional machine learning methods, such as SVM and MLP, are computationally fast, generally taking less than 0.1 ms on average, yet the accuracies of corrosion identification are less than 0.9.

4.2.3. Effect of distributed sensor parameters

The effects of three important sensor parameters on the accuracy of the proposed machine learning model have been evaluated. The investigated sensor parameters are the spatial resolution, coating thickness, and helix spacing, as shown in Table 4. Spatial resolution is an important sensing parameter which describes the spatial fineness of the measurements from DFOS. Coating thickness is associated with the type of fiber optic cable, and it affects the strain transfer behavior [36] and mechanical properties of DFOS. Helix spacing is associated with the efficiency of sensor installation, and coarse spacing is preferred for achieving a high installation efficiency when the accuracy of measurement is sufficient. The investigated spatial resolutions are 0.65 mm, 1.30 mm, 2.60 mm, and 5.20 mm, which are consistent with the OFDR technology. The investigated coating thicknesses are 242 μm , 650 μm , and 900 μm , which represent typical commercial fiber optic cables. The investigated helix spacings are 10 mm, 20 mm, 40 mm, 60 mm, and 80 mm, consistent with the literature [10]. The results show that the proposed machine learning model successfully detects corrosion in different

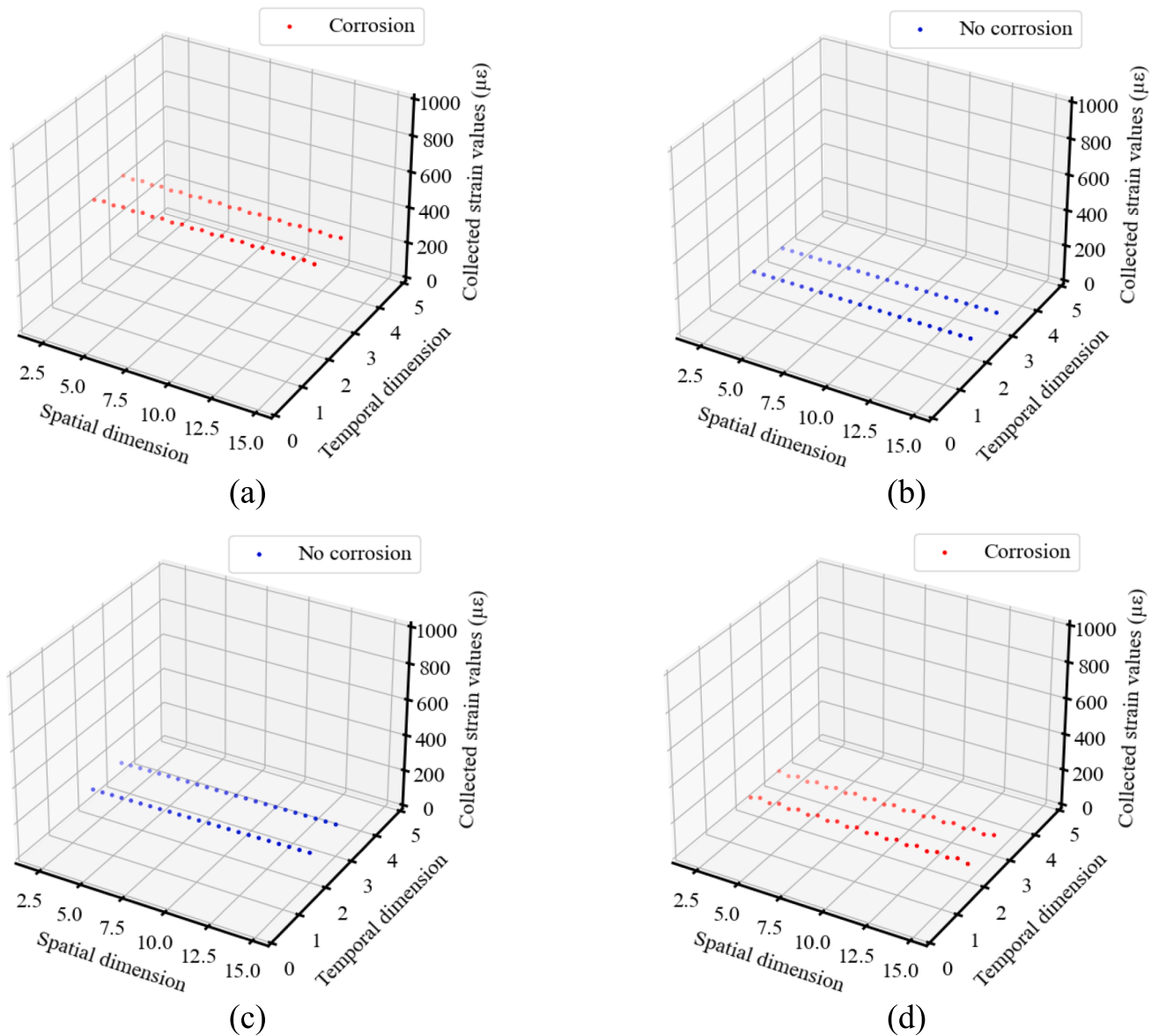


Fig. 13. Examples of corrosion detection results using the proposed machine learning model: (a) TP, (b) TN, (c) FN, and (d) FP.

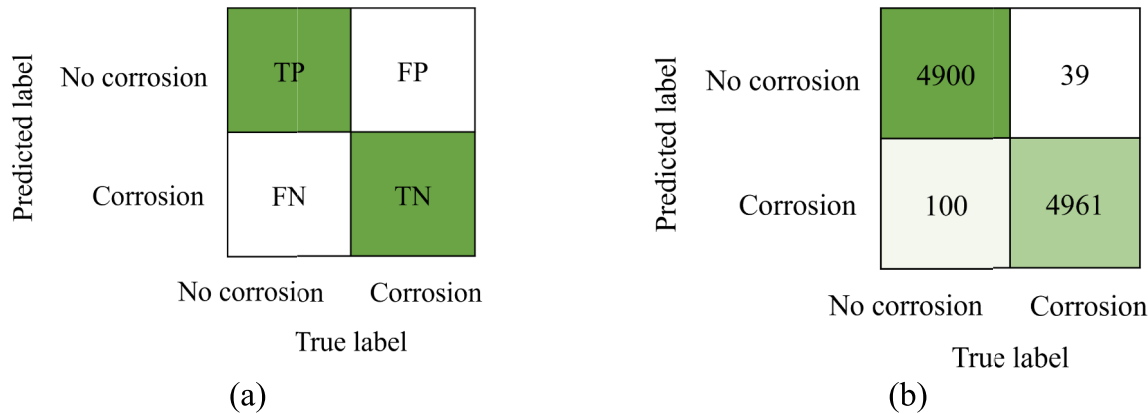


Fig. 14. Confusion matrix of the proposed machine learning model: (a) illustration of the matrix, and (b) the specific matrix in this study.

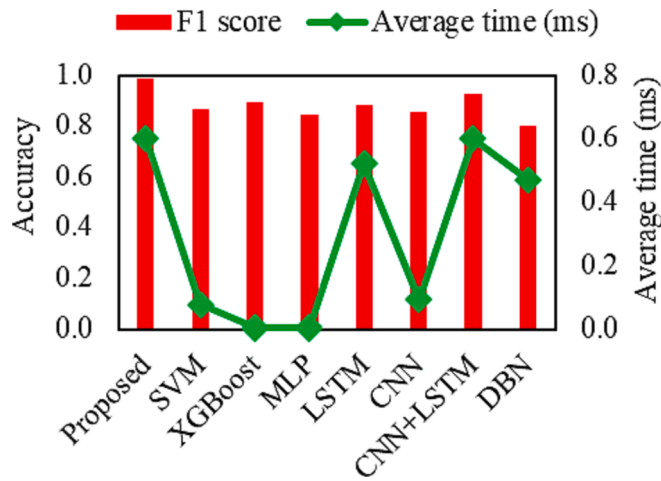


Fig. 15. Performance comparison for the proposed machine learning model and other models.

Table 4

Effect of the parameters of the distributed sensors.

No.	Spatial resolution (mm)	Coating thickness (μm)	Spacing length (mm)	F1 score
1	0.65	900	20	0.977
2	1.30	900	20	0.968
3	2.60	900	20	0.963
4	5.20	900	20	0.950
5	0.65	242	20	0.980
6	0.65	650	20	0.977
7	0.65	242	10	0.982
8	0.65	242	40	0.977
9	0.65	242	60	0.971
10	0.65	242	80	0.979

scenarios with F1 scores higher than 0.95, revealing that the proposed approach is effective in various cases involving different sensor parameters.

4.2.4. Robustness

The robustness of the proposed approach was evaluated using six new datasets. The first dataset was collected from a similar corrosion experiment using large low-carbon steel pipes with a diameter of 38.1 mm, as shown in Fig. 8(c). The experiment with the large steel pipe increases the variability of the dataset compared to the previous experiments, which allows for a better simulation of the pipe conditions in real-world scenarios, which in turn comes to validate the robustness of the proposed method. The remaining five datasets were produced by adding Gaussian white noises to the data obtained from the corrosion experiment using small low-carbon steel pipes. Various noise-to-signal ratios (1 %, 2 %, 3 %, 4 %, and 5 %) of Gaussian white noise were included in the data. The percentages indicate the amplitude of the strain data. The results of the proposed approach are shown in Table 5.

The baseline performance was established using the test data from

Table 5

Robustness of the proposed approach on corrosion detection.

No.	Source of the dataset	Pipe diameter (mm)	F1 score
/	Baseline	25.4	0.986
1	Experiments using large pipes	38.1	0.975
2	Baseline + 1 % noise	25.4	0.981
3	Baseline + 2 % noise	25.4	0.985
4	Baseline + 3 % noise	25.4	0.984
5	Baseline + 4 % noise	25.4	0.984
6	Baseline + 5 % noise	25.4	0.982

the corrosion experiment of the small low-carbon steel pipes. The F1 score achieved by the proposed method on the baseline dataset is 0.986. On the larger low-carbon steel pipes, the F1 score slightly decreases to 0.975. Despite this minor reduction, comparative experiments highlight the commendable robustness of the proposed method across various corrosion datasets. When a larger, diverse dataset becomes available in the future, the proposed method can be applied to train and validate a model with better performance in terms of accuracy and generalizability.

4.3. Quantification results and discussion

4.3.1. Quantification results

According to Refs. [10,37], the volume expansion coefficient (λ) of rust is approximately 2.0 for steel. The results of mass loss obtained from the proposed approach and corrosion tests are compared in Fig. 16.

An approach proposed in Ref. [10] was employed for comparison with the proposed approach. The green line shows the true mass loss obtained from corrosion tests. The R^2 value of the proposed approach is 0.953, and the R^2 value of the reference approach is 0.656. The results indicate that the proposed approach can accurately quantify corrosion severity.

4.3.2. Effect of distributed sensor parameters

The effects of spatial resolution, coating thickness, and helix spacing on the quantification accuracy of the proposed machine learning model have been evaluated, as shown in Fig. 17. As the spatial resolution and coating thickness change, the errors are relatively stable, meaning that the quantification accuracy is insensitive to the spatial resolution and coating thickness. As the helix spacing increases from 10 mm to 80 mm, the error is increased from 0.231 g to 0.384 g, revealing that fine spacing is preferred for improving the corrosion quantification accuracy. Since fine spacing compromises the sensor installation efficiency, it is recommended to adopt helix spacings between 40 mm and 60 mm.

4.4. Limitations and future opportunities

In this research, the method for automatic interpretation of DFOS data for pipeline corrosion was developed based on laboratory

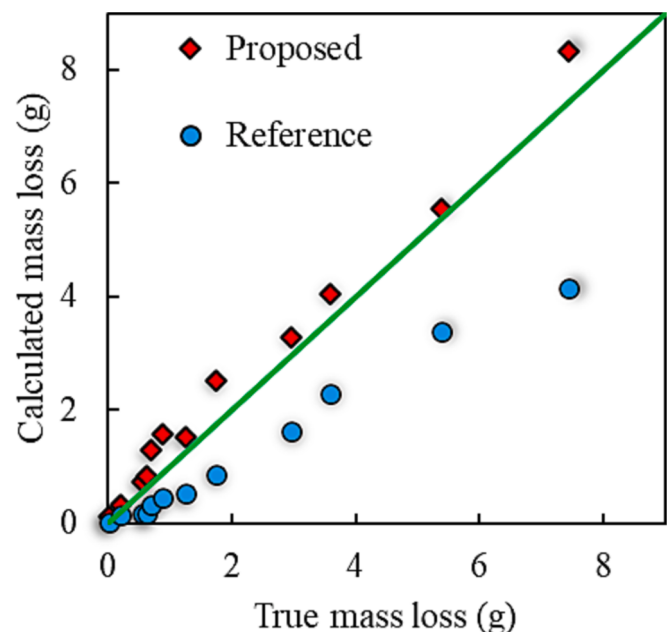


Fig. 16. Corrosion quantification results of the proposed and reference approaches [10].

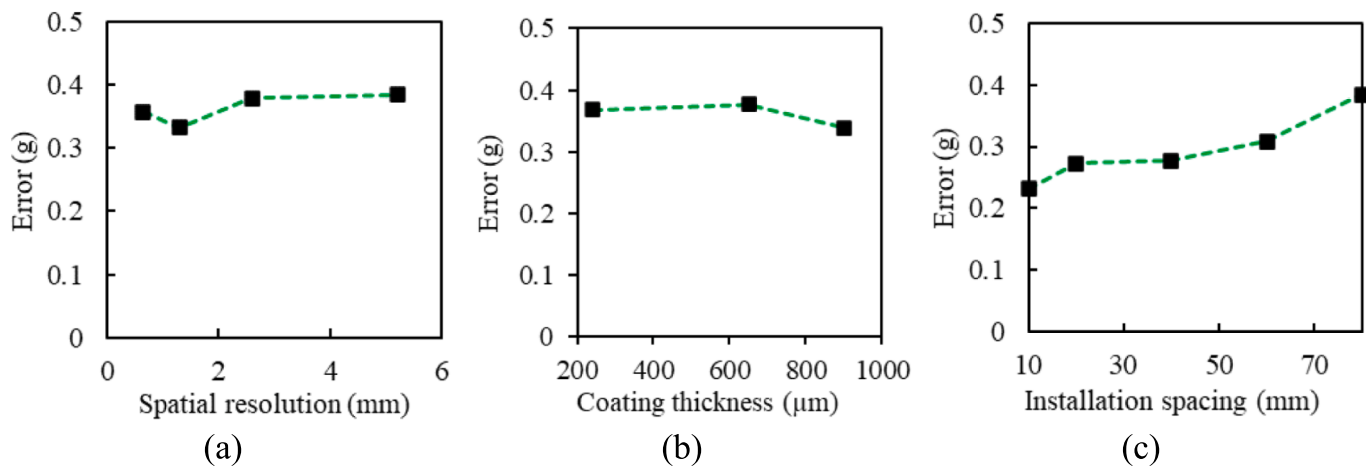


Fig. 17. Effects of sensor parameters on corrosion quantification accuracy: (a) spatial resolution, (b) coating thickness, and (c) helix spacing.

experiments. It should be noted that real applications will involve more variables, such as the variations of temperature and environmental vibrations, which will potentially influence the performance of the proposed method. The main limitation of this research is that the dataset utilized to train and test the performance of the machine learning model is limited, in terms of the size and diversity. Currently, this is a general problem for using machine learning to automate the interpretation of DFOS data because there is lack of large datasets for DFOS data. In particular, there is lack of datasets for field testing data because DFOS has not been widely used in engineering structures. When DFOS is applied to more projects, more diverse data will be generated. However, the lack of effective methods for automatic data interpretation is one of the important challenges for the adoption of DFOS in real applications because DFOS generates a large number of data which cause difficulties in data analysis when data analysis is performed manually. For example, when the spacing between adjacent measuring points is 0.65 mm, each measurement from a DFOS with a length of 100 m generates 153,800 data points. When the sampling frequency is 20 Hz, more than 3 million data points will be generated every second. Although DFOS can provide real-time measurements, the measurement data cannot be manually analyzed and interpreted in real time. More laboratory and field tests are necessary to fully evaluate and improve the performance of the proposed method.

In future research, it is promising to develop knowledge-guided machine learning methods which combine the strengths of established principles and data-driven methods. For example, the loss function of the machine learning models can include two terms, one term for the compliance with established principles such as governing equations and the other term for the compliance with data from experiments. In the training process, the machine learning models will be trained to satisfy both established principles and the data. It is speculated that high accuracy and generalizability will be achieved using knowledge-guided machine learning methods even when the dataset is small. This speculation needs to be tested through further research. In addition to the development of knowledge-guided machine learning methods, it is important to conduct more laboratory and field experiments to establish a larger database, which will be significant for training and testing machine learning models for DFOS sensor data analysis and interpretation.

5. Conclusions

This paper proposes an automated corrosion monitoring approach using strain data measured from DFOSs and machine learning. The following conclusions are drawn:

- The proposed machine learning model is successfully developed to automatically detect corrosion using the DFOS data. Bayesian optimization based hyperparameter tuning was used to find the hyperparameter combinations that performed well from millions of hyperparameter candidate combinations to improve the detection accuracy of the proposed machine learning model. The F1 score of the proposed model increased from 0.835 to 0.986, revealing a high accuracy of the machine learning model for corrosion detection and the effectiveness of the Bayesian optimization based hyperparameter tuning.
- The robustness of the proposed approach in corrosion detection is analyzed. High values (>0.950) of the F1 score were retained, revealing that the proposed approach is robust to the different pipe sizes, environmental noises, and the parameters of the DFOS, including different coating thickness, spatial resolution, and installation spacing of the DFOS.
- Based on the results of the machine learning model, a corrosion quantification method is proposed for both uniform and non-uniform cases. The proposed quantification method considers that some of the measured strains are not generated by corrosion. The R^2 of the proposed quantification method reached 0.953, while that of the reference method was 0.656, indicating a significant improvement.
- The effects of the coating thickness, spatial resolution, and installation spacing of the DFOS on the accuracy of the proposed corrosion quantification method were investigated. The coating thickness and the spatial resolution of the DFOS do not have a significant effect on the accuracy. However, the larger the winding spacing, the smaller the accuracy of the quantification method. When the winding spacing increased from 10 mm to 80 mm, the error of the mass loss due to corrosion increased from 0.231 g to 0.384 g.

The validation and performance evaluation are limited to laboratory experiments in this study. It should be noted that more variables can be involved in field applications, such as variations in temperature and environmental vibrations, which were not included in the laboratory experiments. Before this method is implemented into real projects, it is essential to validate and enhance the performance using larger, more diverse datasets, as well as field testing using in-use pipelines.

CRedit authorship contribution statement

Yiming Liu: Data curation, Formal analysis, Investigation, Software, Visualization, Writing – original draft. **Xiao Tan:** Data curation, Investigation, Validation, Writing – review & editing. **Yi Bao:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yi Bao reports financial support was provided by the U.S. Department of Transportation and the U.S. National Science Foundation.

Data availability

Data will be made available on request.

Acknowledgement

This research was funded by the U.S. Department of Transportation through PHMSA [grant number: 693JK31950008CAAP and 693JK32310008POTA] and the U.S. National Science Foundation [grant number CPS 2305882].

References

- J.Y. Hu, S.S. Zhang, E. Chen, W.G. Li, A review on corrosion detection and protection of existing reinforced concrete (RC) structures, *Constr. Build. Mater.* 325 (2022) 126718, <https://doi.org/10.1016/j.conbuildmat.2022.126718>.
- G.A. Jacobson, NACE international's IMPACT study breaks new ground in corrosion management research and practice, *The Bridge* 46 (2016). <https://www.nae.edu/Publications/Bridge/155266/155346.aspx>.
- M. Vitali, C. Zuliani, F. Corvaro, B. Marchetti, F. Tallone, Statistical analysis of incidents on onshore CO2 pipelines based on PHMSA database, *J. Loss Prev. Process Ind.* 77 (2022) 104799, <https://doi.org/10.1016/j.jlp.2022.104799>.
- Y. Liu, Y. Bao, Review on automated condition assessment of pipelines with machine learning, *Adv. Eng. Inf.* 53 (2022) 101687, <https://doi.org/10.1016/j.aei.2022.101687>.
- S. Esmailzadeh, M. Aliofkhaeaei, H. Sarlak, Interpretation of cyclic potentiodynamic polarization test results for study of corrosion behavior of metals: A review, *Prot. Met. Phys. Chem* 54 (2018) 976–989, <https://doi.org/10.1134/S207020511805026X>.
- L. Chen, R.K.L. Su, Corrosion rate measurement by using polarization resistance method for microcell and macrocell corrosion: Theoretical analysis and experimental work with simulated concrete pore solution, *Constr. Build. Mater.* 267 (2021) 121003, <https://doi.org/10.1016/j.conbuildmat.2020.121003>.
- S. Feliu, Electrochemical impedance spectroscopy for the measurement of the corrosion rate of magnesium alloys: Brief review and challenges, *Metals* 10 (2020) 1–23, <https://doi.org/10.3390/met10060775>.
- R.F. Wright, P. Lu, J. Devkota, F. Lu, M. Ziomek-Moroz, P.R. Ohodnicki, Corrosion sensors for structural health monitoring of oil and natural gas infrastructure: A review, *Sensors* 19 (2019) 3964, <https://doi.org/10.3390/s19183964>.
- X. Peng, U. Anyaoha, Z. Liu, K. Tsukada, Analysis of magnetic-flux leakage (MFL) data for pipeline corrosion assessment, *IEEE Trans. Magn.* 56 (2020) 1–15, <https://doi.org/10.1109/TMAG.2020.2981450>.
- X. Tan, L. Fan, Y. Huang, Y. Bao, Detection, visualization, quantification, and warning of pipe corrosion using distributed fiber optic sensors, *Autom. Constr.* 132 (2021) 103953, <https://doi.org/10.1016/j.autcon.2021.103953>.
- H. Ma, B. Zhao, Z. Liu, C. Du, B. Shou, Local chemistry–electrochemistry and stress corrosion susceptibility of X80 steel below disbonded coating in acidic soil environment under cathodic protection, *Constr. Build. Mater.* 243 (2020) 118203, <https://doi.org/10.1016/j.conbuildmat.2020.118203>.
- A. Nishikata, Q. Zhu, E. Tada, Long-term monitoring of atmospheric corrosion at weathering steel bridges by an electrochemical impedance method, *Corros. Sci.* 87 (2014) 80–88, <https://doi.org/10.1016/j.corsci.2014.06.007>.
- V. Marcantonio, D. Monarca, A. Colantoni, M. Cecchini, Ultrasonic waves for materials evaluation in fatigue, thermal and corrosion damage: A review, *Mech. Syst. Sig. Process.* 120 (2019) 32–42, <https://doi.org/10.1016/j.ymssp.2018.10.012>.
- L. Fan, X. Shi, Techniques of corrosion monitoring of steel rebar in reinforced concrete structures: A review, *Struct. Health Monit.* 21 (2022) 1879–1905, <https://doi.org/10.1177/14759217211030911>.
- M. Zeng, H. Chen, J. Ling, H. Zhao, D. Wu, Monitoring of prestressing forces in cross-tensioned concrete pavements during construction and maintenance based on distributed optical fiber sensing, *Autom. Constr.* 142 (2022) 104526, <https://doi.org/10.1016/j.autcon.2022.104526>.
- M. Li, X. Feng, Y. Han, Brillouin fiber optic sensors and mobile augmented reality-based digital twins for quantitative safety assessment of underground pipelines, *Autom. Constr.* 144 (2022) 104617, <https://doi.org/10.1016/j.autcon.2022.104617>.
- T. Jiang, L. Ren, Z.G. Jia, D.S. Li, H.N. Li, Pipeline internal corrosion monitoring based on distributed strain measurement technique, *Struct. Control Health Monit.* 24 (2017) 1–11, <https://doi.org/10.1002/stc.2016>.
- W. Shen, Q. Pang, L. Fan, P. Li, X. Zhao, Monitoring and quantification of non-uniform corrosion induced mass loss of steel piles with distributed optical fiber sensors, *Autom. Constr.* 148 (2023) 104769, <https://doi.org/10.1016/j.autcon.2023.104769>.
- Q. Song, Y. Chen, E. Abdoli Oskoui, Z. Fang, T. Taylor, G. Tang, X. Zhao, F. Ansari, Micro-crack detection method of steel beam surface using stacked autoencoders on massive full-scale sensing strains, *Struct. Health Monit.* 19 (2020) 1175–1187, <https://doi.org/10.1177/1475921719879965>.
- Q. Song, G. Yan, G. Tang, F. Ansari, Robust principal component analysis and support vector machine for detection of microcracks with distributed optical fiber sensors, *Mech. Syst. Sig. Process.* 146 (2021) 107019, <https://doi.org/10.1016/j.ymssp.2020.107019>.
- Y. Liu, Y. Bao, Automatic interpretation of strain distributions measured from distributed fiber optic sensors for crack monitoring, *Measurement* 211 (2023) 112629, <https://doi.org/10.1016/j.measurement.2023.112629>.
- Y. Liu, Y. Bao, Intelligent monitoring of spatially-distributed cracks using distributed fiber optic sensors assisted by deep learning, *Measurement* 220 (2023) 113418, <https://doi.org/10.1016/j.measurement.2023.113418>.
- I. The Corning, Corning® SMF-28e+® optical fiber production information, 2023. <https://www.corning.com/media/worldwide/coc/documents/Fiber/PI-146-3-AEN.pdf> (accessed May 1, 2023).
- S. Yeoh, M.Z. Matjafri, K.N. Mutter, A.A. Oglat, Plastic fiber evanescent sensor in measurement of turbidity, *Sensors Actuators, a: Physical* 285 (2019) 1–7, <https://doi.org/10.1016/j.sna.2018.10.042>.
- LUNA, 2022. ODISI 6000 series: Optical distributed sensor interrogators data sheet. pp. 1–6. [https://lunainc.com/sites/default/files/assets/files/data-sheet/Luna ODISI 6000 Data Sheet.pdf](https://lunainc.com/sites/default/files/assets/files/data-sheet/Luna%20ODISI%206000%20Data%20Sheet.pdf) (accessed July 1, 2023).
- J. Wang, K. Yu, J. Bu, Y. Lin, S. Han, Multi-classification of UWB signal propagation channels based on one-dimensional wavelet packet analysis and CNN, *IEEE Trans. Veh. Technol.* 71 (2022) 8534–8547, <https://doi.org/10.1109/tvt.2022.3172863>.
- Y. Liu, Y. Bao, Real-time remote measurement of distance using ultra-wideband (UWB) sensors, *Autom. Constr.* 150 (2023) 104849, <https://doi.org/10.1016/j.autcon.2023.104849>.
- J.S. Chou, T.K. Nguyen, Forward Forecast of Stock Price Using Sliding-Window Metaheuristic-Optimized Machine-Learning Regression, *IEEE Trans. Ind. Inf.* 14 (2018) 3132–3142, <https://doi.org/10.1109/TII.2018.2794389>.
- S. Shu, L. Zhang, Y. Sun, J. Tang, Host-Parasite: Graph LSTM-in-LSTM for group activity recognition, *IEEE Trans. Neural Networks Learn. Syst.* 32 (2021) 663–674, <https://doi.org/10.1109/TNNLS.2020.2978942>.
- B. Wu, Z. Wang, K. Chen, C. Yan, W. Liu, GBC: An energy-efficient LSTM accelerator with gating units level balanced compression strategy, *IEEE Trans. Circuits Syst. I Regul. Pap.* 69 (2022) 3655–3665, <https://doi.org/10.1109/TCSI.2022.3181975>.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need. *Advances in Neural Information Processing Systems*, 2017–Decem, 2017, pp. 5999–6009.
- D. Soydaner, Attention mechanism in neural networks: where it comes and where it goes, *Neural Comput. Applic.* 34 (2022) 13371–13385, <https://doi.org/10.1007/s00521-022-07366-3>.
- Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, *Neurocomputing* 452 (2021) 48–62, <https://doi.org/10.1016/j.neucom.2021.03.091>.
- X. Yin, Y. Chen, A. Bouferguene, H. Zaman, M. Al-Hussein, L. Kurach, A deep learning-based framework for an automated defect detection system for sewer pipes, *Autom. Constr.* 109 (2020) 102967, <https://doi.org/10.1016/j.autcon.2019.102967>.
- V.D. Earshia, M. Sumathi, A comprehensive study of 1D and 2D image interpolation techniques, in: *Proceedings of the International Conference on Communications and Cyber Physical Engineering* 2018, vol. 500, 2019, pp. 497–511. <https://doi.org/10.1007/978-981-13-0212-1>.
- X. Tan, Y. Bao, Q.H. Zhang, H. Nassif, G. Chen, Strain transfer effect in distributed fiber optic sensors under an arbitrary field, *Autom. Constr.* 124 (2021) 103597, <https://doi.org/10.1016/j.autcon.2021.103597>.
- X. Zhao, P. Gong, G. Qiao, J. Lu, X. Lv, J. Ou, Brillouin corrosion expansion sensors for steel reinforced concrete structures using a fiber optic coil winding method, *Sensors* 11 (2011) 10798–10819, <https://doi.org/10.3390/s111110798>.