Wiley
Structural Control and Health Monitoring
Volume 2024, Article ID 2337057, 19 pages
https://doi.org/10.1155/2024/2337057

WILEY

Research Article

Behavior Expectation-Based Anomaly Detection in Bridge Deflection Using AOA-BiLSTM-TPA: Considering Temperature and Traffic-Induced Temporal Patterns

Guang Qu,[1] **Ye Xia**,[1,2,3] **Limin Sun**,[1,2,3] **and Gongfeng Xin**[4]

[1]*Department of Bridge Engineering, School of Civil Engineering, Tongji University, Siping Rd 1239, Shanghai 200092, China*
[2]*State Key Laboratory for Disaster Reduction in Civil Engineering, Tongji University, Siping Rd 1239, Shanghai, China*
[3]*Shanghai Qi Zhi Institute, Yunjing Road 701, Xuhui, Shanghai 200232, China*
[4]*Innovation Research Institute of Shandong High-Speed Group Co., LTD, Zhoucun, China*

Correspondence should be addressed to Ye Xia; yxia@tongji.edu.cn

Received 6 March 2024; Revised 28 April 2024; Accepted 16 May 2024; Published 1 June 2024

Academic Editor: jian zhang

In the realm of structural health monitoring (SHM), understanding the expected behavior of a structure is vital for the timely identification of anomalous activities. Existing methods often model only the physical quantities of monitoring data, neglecting the corresponding temporal information. To address this, this paper presents an innovative deep learning framework that synergistically combines a BiLSTM model, fortified by a temporal pattern attention (TPA) mechanism, with time-encoded temperature and traffic-induced deflection-temporal patterns. The arithmetic optimization algorithm (AOA) is employed for optimal hyperparameter tuning, and incremental learning was implemented to enable real-time updates of the model. Based on the proposed framework, an anomaly detection method was subsequently developed. This method is bidirectional: it uses quantile loss to provide expected ranges for structural behavior, identifying isolated anomalies, while the windowed normalized mutual information (WNMI) based on multivariate kernel density estimation (MKDE) helps detect trend variability caused by decreases in structural stiffness. This framework and the anomaly detection method were validated using data from an operational cable-stayed bridge. The results demonstrate that the method effectively predicts structural behavior and detects anomalies, highlighting the critical role of temporal information in SHM.

1. Introduction

Given the large number of bridges and limited human and financial resources, continuously inspecting every bridge with traditional methods is impractical [1–3]. Currently, inspections occur every 2 to 6 years as per regulatory guidelines [4], but damage occurring between inspections can go unnoticed, leading to further deterioration and increased repair costs. Consequently, it is essential to adopt advanced structural health monitoring (SHM) technologies for real-time anomaly detection in bridges [5, 6]. Bridge deflection is a crucial indicator of structural stiffness and safety [7], affected by various uncertainties such as weather and traffic density [8]. The impact of natural factors, primarily temperature variations, is relatively straightforward to understand [9], but human-induced factors such as traffic flow are less predictable [10]. Gaining insights into these patterns can provide a more comprehensive understanding of a bridge's expected behavior. If an urban bridge consistently shows deflections during peak hours that do not revert to baseline, this suggests potential structural problems, especially if similar patterns have historically led to failures [11–13].

In recent years, various methods have been explored for abnormality diagnosis using bridge SHM data. For example, by integrating digital twin technology with the finite element method (FEM), Jasiński et al. [7] provided a case study of an actual bridge, demonstrated how to generate computational FEM models using building information modeling (BIM), and delved into the results of load tests conducted on the

bridge. The core concept is to employ the finite element model as a digital twin to supplement monitoring systems and simulate scenarios that might initially be interpreted as structural damage [4]. Numerical models have been developed to cross-verify data from sensors [14]. In reality, characterizing expected structural behavior under standard conditions becomes challenging without numerical simulations, making it difficult to detect anomalies [15]. From a generalizability standpoint, these methods may not be entirely practical. Moreover, the models are often constrained by the limitations inherent to the materials and physics involved, hindering their ability to closely mimic real-world operational conditions [16, 17].

Another avenue that offers relative simplicity and efficiency lies in data-driven approaches, particularly leveraging the rapid advancements in machine learning and deep learning techniques. For instance, Nguyen and Goulet [18] have combined the Rao–Blackwellized particle filter with the Bayesian dynamic linear model to create an online method for the real-time detection of abnormal dam displacements. Wang et al. [19] used sparse Bayesian learning to estimate the expected structural behavior under normal conditions, using this as a baseline to categorize damage with a constructed structural damage index. Yue et al. [20] established a digital regression model using long short-term memory (LSTM) networks to accurately map the relationship between temperature features and temperature-induced deflection in the main girder of cable-stayed bridges. Manzini et al. [21] investigated the potential of low-cost global navigation satellite systems (GNSS) for large-scale bridge monitoring, introducing an anomaly detection strategy based on machine learning models, including recurrent neural networks. Zhang et al. [22] introduced a data anomaly detection method based on structural vibration signals and convolutional neural networks (CNN) aimed at automatically identifying and eliminating abnormal data in structural monitoring systems, which proved effective in detecting various abnormal patterns within large datasets generated by long-span bridge monitoring. Additionally, advanced methods such as support vector regression (SVR) [23–25], gated recurrent units (GRU) [26], and transform coupled with a random forest classifier [27] have been used for anomaly detection in SHM data. A review by Yue et al. [28] highlights the latest applications of deep learning in SHM, particularly emphasizing the capabilities of LSTM networks in time-series modeling. In SHM systems, the response data traditionally collected is often viewed as only one-dimensional, focusing solely on the response itself. However, an often overlooked aspect is the timestamp associated with this monitoring data. For instance, a 24-hour timestamp can provide crucial periodic information, and peak traffic periods can also be highlighted temporarily. Existing methods typically fail to capture these subtle temporal differences and do not effectively utilize the valuable information present in the time dimension. Therefore, there is an urgent need to develop an anomaly detection framework that can effectively utilize overlooked temporal information, thereby enhancing the performance of SHM systems.

This paper proposes an arithmetic optimization algorithm-based bidirectional long short-term memory with temporal pattern attention (AOA-BiLSTM-TPA) for predicting bridge deflection and utilizes the expected deflection behavior for anomaly detection. To fully leverage temporal information, this paper implements two types of encoding on the monitoring timeline. Cyclic encoding enhances the BiLSTM's ability to capture temperature-induced deflection patterns, while label encoding improves the learning of traffic-induced deflections. Additionally, TPA is used to assign weights to these patterns. To facilitate real-time anomaly detection in measurement data, this method uses incremental learning to continuously update the model. It utilizes quantile loss to provide safety margin thresholds and incorporates window-normalized mutual information (WNMI) to check the consistency between expected and actual trends.

2. Encoding Techniques for Temporal Influences on Structural Deflections

In this paper, hourly-averaged deflection is used as a research metric to diminish noise impacts and reduce the misleading effects of transient responses on anomaly detection. Nonetheless, such deflections are subject to complex, time-variant factors, including temperature changes and traffic patterns. To fully leverage temporal information within deep learning methodologies, cyclic encoding has been utilized to enhance the model's ability to learn patterns of temperature-induced deflections. Simultaneously, label encoding has been used to improve the learning of patterns associated with traffic-induced deflections.

2.1. Cyclic Encoding for Temperature-Induced Deflections. The cyclical pattern of temperature-induced deflections follows a clear 24-hour cycle [29]. Linear encoding, which assigns each hour a unique value from 0 to 23, inadequately reflects this time's cyclical essence, possibly leading to imprecise forecasts or less-than-ideal outcomes in machine learning or optimization. To overcome this, our study adopts a circular encoding approach. This method maps the data onto a two-dimensional plane as follows:

$$\begin{aligned} \mathbf{X}_{\sin}(t) &= \sin\left(\frac{2\pi t}{24}\right), \\ \mathbf{X}_{\cos}(t) &= \cos\left(\frac{2\pi t}{24}\right), \end{aligned} \tag{1}$$

where t represents the timestamp. The points $(\mathbf{X}_{\sin}(t), \mathbf{X}_{\cos}(t))$ are evenly distributed on the unit circle, creating a unique 2D representation of the time variable that captures the circular nature of the data. The cyclic encoding of temporal information helps provide deep learning with periodic insights similar to those of temperature, enhancing interpretability, and can be seen in Figure 1:

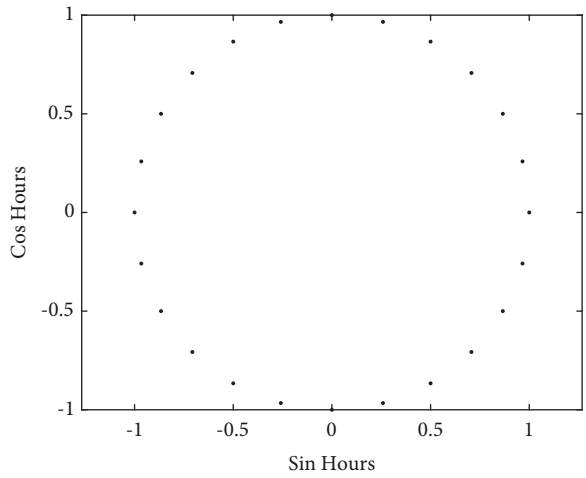

Figure 1: Cyclic encoding (24 hours).

2.2. Label Encoding for Traffic-Induced Deflections. Considering that this study focuses on the hourly-averaged deflection index, the literature [30–32] indicates that traffic congestion during peak periods significantly increases the static deflection of bridges. To fully leverage the temporal information associated with this traffic-induced deflection pattern, this paper uses a dual-layer coding method: the first layer distinguishes between workdays and nonworkdays, while the second layer differentiates between rush hours and nonrush hours. Let W_t represent the first-layer coding for distinguishing workdays by hour and R_t denote the second-layer coding for identifying rush hours at the same hour. These variables can be defined as follows:

$$W_t = \begin{cases} 0, & \text{workday}, \\ 1, & \text{non} - \text{workday}, \end{cases} \quad R_t = \begin{cases} 0, & \text{rush hour}, \\ 1, & \text{non} - \text{rush hour}. \end{cases} \tag{2}$$

In this paper, the operational conditions of the bridge being studied dictate that rush hours are defined as workday morning rush hours (Beijing time 07:00–09:00) and evening rush hours (Beijing time 17:00–19:00). Using the dual-layer coding method, the hourly-averaged deflection at any given moment can be represented by encoding $[W_t \ R_t]$. For instance, the hourly-averaged deflection from 17:00–18:00 on April 20, 2020 (Monday) would be encoded as [1, 1]. This approach, incorporating time labels related to traffic conditions, can provide richer information and enhance interpretability for deep learning.

3. Methodology for Anticipated Behavior Estimation

This section introduces a framework designed to predict deflection behavior. It integrates BiLSTM [33–35] and TPA mechanisms [36] within an AOA-based architecture [37], enhancing the capture of temporal patterns and prediction precision.

3.1. BiLSTM. The foundational LSTM has a unique architecture with three key gates: input, forget, and output. This structure enables the capture of long-term dependencies, surpassing the capabilities of standard RNNs [38]. The forget gate f_t, the input gate i_t, and the output gate o_t can be calculated with

$$\begin{aligned} f_t &= \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right), \\ i_t &= \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right), \\ o_t &= \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right). \end{aligned} \tag{3}$$

In the above equations, $\sigma(\cdot)$ is the sigmoid function, and the input to the gate is a combination of the current input x_t and the previous hidden state h_{t-1}, which are linearly transformed using weights (W_f, W_i, and W_o) and bias terms (b_f, b_i, and b_o).

Candidate memory $\tilde{c}_t$ was generated by applying the hyperbolic tangent (tanh) activation function to a linear transformation using the weight W_c and the bias b_c:

$$\tilde{c}_t = \tanh\left(W_c \cdot \left[h_{t-1}, x_t\right] + b_c\right). \tag{4}$$

The memory cell updated state for the current time step amalgamates c_{t-1} and $\tilde{c}_t$ and is influenced by the decisions of both the forget and input gates:

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t. \tag{5}$$

The hidden state is the output of the LSTM at the current step:

$$h_t = o_t \cdot \tanh\left(c_t\right). \tag{6}$$

The BiLSTM extends LSTM by processing sequences in both forward and reverse directions [35]. The final prediction combines hidden states from both directions:

$$h_t = \overrightarrow{h}_t \oplus \overleftarrow{h}_t, \tag{7}$$

with h_t encapsulating the forward $\overrightarrow{h}_t$ and backward $\overleftarrow{h}_t$ directions and $\oplus$ denoting the concatenation operation. The resultant h_t is derived by merging the forward and backward hidden states at time t. The detailed schematic of the BiLSTM is shown in Figure 2:

3.2. TPA Mechanism. In this paper, TPA is utilized to capture the temporal dependencies in the monitoring data. In the TPA framework, convolutional neural networks (CNNs) are employed to scan the input time-series data using a set of k convolutional filters $C_i \in \mathbb{R}^{1\times T}$, where $T = w$ (if unspecified) is the maximum sequence length considered. The convolution operation is defined as follows:

$$H_{i,j}^{C} = \sum_{l=1}^{w} H_{i,(t-w-1+l)} \times C_{j,T-w+l}, \tag{8}$$

where $H_{i,j}^{C}$ represents the convolutional output, capturing localized temporal patterns in each of its row vectors, and $H^{C} \in \mathbb{R}^{n\times k}$.

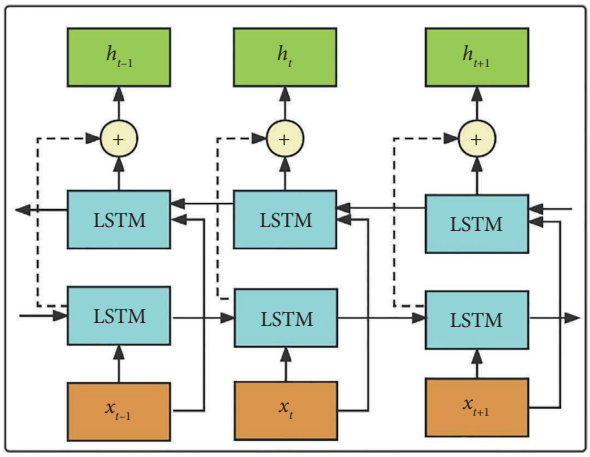

FIGURE 2: BiLSTM diagram.

The attention mechanism uses a scoring function to weigh the importance of each feature, which is defined as $f: \mathbb{R}^k \times \mathbb{R}^m \mapsto \mathbb{R}$ to evaluate relevance:

$$f\left(H_i^C, h_t\right) = \left(H_i^C\right)^{\mathrm{T}} W_a h_t, \tag{9}$$

where H_i^C is the ith row of H^C, and $W_a \in \mathbb{R}^{k\times m}$. The attention weight α_i is then calculated using the sigmoid activation function, allowing for multiple features to be considered relevant:

$$\alpha_i = \mathrm{sigmoid}\left(f\left(H_i^C, h_t\right)\right). \tag{10}$$

The vector v_t is obtained as a weighted sum of the rows of H^C:

$$v_t = \sum_{i=1}^{n} \alpha_i H_i^C. \tag{11}$$

Then, the final step integrates the context vector v_t and the original feature vector h_t to make the final prediction.

$$\begin{aligned} h_t' &= W_h h_t + W_v v_t, \\ y_{t-1+\Delta} &= W_{h'} h_t', \end{aligned} \tag{12}$$

3.3. AOA. The optimization of the number of hidden layers in BiLSTMs is a critical aspect that significantly influences model performance. Conventional methods, including manual grid search and traditional optimization algorithms such as genetic algorithms [39] or particle swarm optimization [40], have limitations. These methods are often computationally intensive, or they may lack the capacity for adaptive searching in a high-dimensional hyperparameter space. Against this backdrop, this study adopts an advanced optimization method, the AOA, specifically for calibrating the number of hidden layers in BiLSTMs. The AOA, first developed by Abualigah et al. [37], leverages four fundamental arithmetic operations (addition, subtraction, multiplication, and division) for optimization tasks. The algorithm can be broken down into three primary phases: initialization, exploration, and exploitation.

(1) Initialization: The candidate solution is defined as

$$y(t) = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,n} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,n} \\ \vdots & \vdots & & \vdots \\ y_{m,1} & y_{m,2} & \cdots & y_{m,n} \end{bmatrix}. \tag{13}$$

In this paper, $n = 10$ represents the number of initial solutions, and $m = 1$ indicates one hyperparameter to be optimized: the number of hidden units.

(2) Exploration: After initialization in the AOA, it first enters the exploration stage. Before this phase, the math optimizer accelerated (MOA) is computed for exploration purposes, which is derived from

$$\mathrm{MOA}\left(C_{\mathrm{iter}}\right) = \mathrm{Min} + C_{\mathrm{iter}} \times \left(\frac{\mathrm{Max} - \mathrm{Min}}{\mathrm{M}_{\mathrm{iter}}}\right), \tag{14}$$

where C_{iter} represents the current iteration, and $C_{\mathrm{iter}} \in (1, M_{\mathrm{iter}})$. *Max* and *Min* indicate the maximum and minimum values of the acceleration function, respectively. $\mathrm{MOA}(C_{\mathrm{iter}})$ denotes the function value at the tth iteration, which is calculated using the following equation:

$$y_{i,j}\left(C_{\mathrm{iter}}+1\right) = \begin{cases} \mathrm{best}\left(y_j\right) \div (\mathrm{MOP}+\varepsilon) \times \left[\left(ub_j - lb_j\right) \times \mu + lb_j\right], & r_2 < 0.5, \\ \mathrm{best}\left(y_j\right) \times \mathrm{MOP} \times \left[\left(ub_j - lb_j\right) \times \mu + lb_j\right], & r_2 \geq 0.5, \end{cases} \tag{15}$$

where $y_{i,j}(C_{\mathrm{iter}}+1)$ denotes the ith solution in the subsequent iteration, $y_{i,j}(C_{\mathrm{iter}}+1)$ signifies the jth position of the ith solution in the present iteration, and $\mathrm{best}(y_j)$ is the jth position in the optimal iteration. ε is a small integer number, while ub_i and ub_i, respectively, represent the upper and lower bounds of the jth position. μ is set to 0.5.

Utilizing multiplication or division as the exploration mechanism of AOA helps in locating a solution as close as possible to the optimum by extensively searching the solution space over multiple iterations. The math optimizer probability, $\mathrm{MOP}(C_{\mathrm{iter}})$, indicating the function value during the t-th iteration, is defined as follows:

$$\mathrm{MOP}\left(C_{\mathrm{iter}}\right) = 1 - \frac{C_{\mathrm{iter}}{}^{(1/\alpha)}}{M_{\mathrm{iter}}{}^{(1/\alpha)}}, \tag{16}$$

where C_{iter} signifies the current iteration, while M_{iter} stands for the maximum iterations. α is a sensitive parameter defining the precision of the exploitation process during iterations. In this study, α is set to 5.

(3) Exploitation: In this phase, subtraction or addition arithmetic operators are utilized to yield high-density results, offering advantages over multiplication and division operators due to their ability to approach the target with minimal dispersion. The exploitation phase is described as follows:

$$y_{i,j}\left(C_{\mathrm{iter}}+1\right) = \begin{cases} \mathrm{best}\left(y_j\right) + \mathrm{MOP} \times \left[\left(ub_j - lb_i\right) \times \mu + lb_j\right], & r_3 \geq 0.5, \\ \mathrm{best}\left(y_j\right) - \mathrm{MOP} \times \left[\left(ub_j - lb_i\right) \times \mu + lb_j\right], & r_3 < 0.5. \end{cases} \tag{17}$$

To effectively balance the exploration and exploitation phases of the algorithm, r_1, a random number between [0, 1], is used in AOA as the criterion to transition from exploration to exploitation. When $r_1 >$ MOA, the candidate solution attempts to diverge from the approximate optimal solution, signifying that the algorithm is in the exploration phase. Conversely, when $r_1 \leq$ MOA, the candidate solution gravitates towards the approximate optimal solution, indicating that the algorithm is in the exploitation phase.

3.4. AOA-BiLSTM-TPA Framework with Incremental Learning. As depicted in Figure 3, the proposed framework consists of several integral components: BiLSTM layers for sequence modeling, CNN convolutional layers for feature extraction, a TPA mechanism for weighted feature integration, and fully connected layers for final output prediction. In this paper, the input length for each timestep of the BiLSTM is set to 24 hours to enhance the mastery of periodic patterns. For the optimization phase, the adaptive moment estimation (Adam) optimizer [41] is used for model optimization, enabling adaptive learning rates for each parameter. The initial learning rate is set at 0.01 and decreases by a predefined factor of 0.9 every five epochs. The training process is limited to a maximum of 200 epochs, with a mini-batch size of 10.

Furthermore, to enhance the real-time applicability of the proposed framework, an incremental learning strategy has been implemented within the neural network. This approach, also known as online learning or continual learning, allows the model to continuously learn and adapt to new data while retaining previously acquired knowledge [42]. In incremental learning, the challenge is to address the problem of catastrophic forgetting, which occurs when a model, upon learning new information, completely disrupts the knowledge it had previously acquired. To address this, this paper uses the experience replay incremental learning strategy [43], which maintains and reuses a memory of past data samples during new data training. The process is as follows:

(1) Historical deflection data are divided into 24-hour subsets to match the input length 24 of BiLSTM in this paper. These subsets are denoted as $\{\mathbf{X}_i\}$, $i = 1, 2, \cdots, n$.

(2) When updating the network state at time t, a subset $\mathbf{X}_i$ is randomly selected from the memory and combined with a new dataset $\tilde{\mathbf{X}}$ that includes all monitoring information (i.e., deflection data and labels) collected from time t-24 to t-1. The training target is set to predict the subsequent value of $\left[\mathbf{X}_i \ \tilde{\mathbf{X}}\right]$.

(3) Training adjustments are minimal, using just 10 epochs and a learning rate of 0.001 to refine the model gently. Each epoch shuffles the data to prevent sequence memorization and enhance generalization.

The incremental learning strategy is shown in Figure 4.

In this research, the root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) are used as evaluation metrics to gauge the reconstruction accuracy of the proposed model. The mathematical representations of these metrics are as follows:

$$\begin{aligned} \mathrm{RMSE} &= \sqrt{\frac{\sum_{t=1}^{n}\left(x_t - \widehat{x}_t\right)^2}{n}}, \\ \mathrm{MAE} &= \frac{1}{n}\sum_{t=1}^{n}\left|x_t - \widehat{x}_t\right|, \\ R^2 &= 1 - \frac{\sum_{t=1}^{n}\left(x_t - \widehat{x}_t\right)^2}{\sum_{t=1}^{n}\left(x_t - \overline{x}\right)^2}. \end{aligned} \tag{18}$$

In the above equations, n represents the sample size, and x_t and $\widehat{x}_t$ are the monitored and predicted data, respectively. A lower RMSE and MAE indicate higher reconstruction accuracy, while an R^2 score closer to 1 suggests a higher fidelity in the reconstructed data to the observed values.

4. Anomaly Detection through AOA-BiLSTM-TPA Model Predictions

Ensuring the structural safety of operational bridges is paramount. Establishing an expected behavior model for the bridge and detecting significant deviations serve as crucial measures for maintaining safety. To enhance the precision and reliability of anomaly detection, this study proposes two

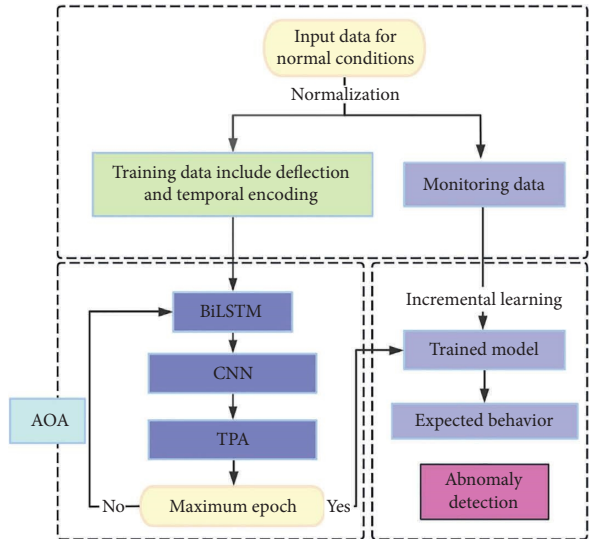

Figure 3: Framework for AOA-BiLSTM-TPA.

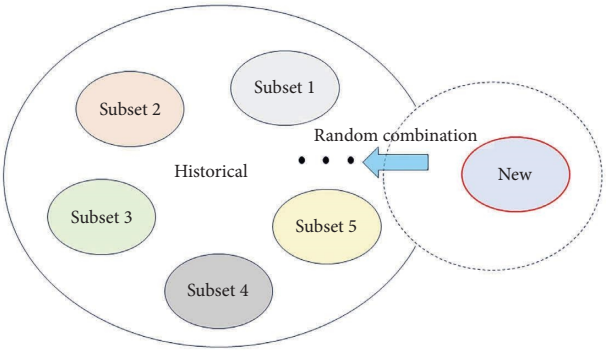

Figure 4: Schematic diagram of the experience replay incremental learning strategy.

strategies. Firstly, it utilizes quantile loss functions during training to identify point anomalies robustly. This method sets quantile threshold intervals, enabling the immediate detection of localized deviations from expected behavior. Secondly, a mutual information metric is used to monitor broader structural behavior trends.

4.1. Loss Functions in Anomaly Detection. Quantile regression is a type of regression analysis used in statistics and econometrics [44]. Unlike standard regression models that focus on the conditional mean of the response variable given certain values of the predictor variables, quantile regression models focus on different quantiles (such as the median or the interquartile range) of the conditional distribution of the response variable. This paper introduces the quantile loss function incorporating the Huber norm [45], smoothing out discontinuities at points where the predicted value equals the true value. This ensures that the loss function is differentiable everywhere:

$$\text{loss}\left(y, y^p\right) = \begin{cases} \dfrac{\left(y^p - y\right)^2}{2\varepsilon}, & \left|y^p - y\right| \le \varepsilon, \\ \left|y^p - y\right| - 0.5\varepsilon, & \left|y^p - y\right| > \varepsilon, \end{cases} \tag{19}$$

where y is the monitored deflection data and y^p is the prediction; q is a quantile; and $\varepsilon > 0$ is a threshold that determines the transition point between the quadratic and linear parts of the Huber function.

4.2. Windowed Normalized Mutual Information for Anomaly Detection. Beyond quantile thresholds, assessing the divergence between anticipated and actual structural behaviors requires advanced statistical metrics. Mutual information (MI) [46], grounded in entropy theory, surpasses conventional correlation measures by capturing complex linear and nonlinear dependencies, offering a comprehensive statistical assessment without presupposing any specific functional relationship between variables.

Let X be the actual measured data and Y be the predicted values based on our AOA-BiLSTM-TPA model. The MI between X and Y can be calculated as follows:

$$I(X,Y) = \iint p(x,y)\log_2\left(\frac{p(x,y)}{p(x)\cdot p(y)}\right)\mathrm{d}x\,\mathrm{d}y, \tag{20}$$

where $I(X,Y)$ is the mutual information; $p(x,y)$ is the joint probability density; and $p(x)$ and $p(y)$ are the marginal probabilities. The MI overcomes the randomness of individual mutual information values, rendering it a deterministic quantity. When using a logarithm base of 2, the unit of mutual information is expressed in bits.

To compute $I(X,Y)$, one needs estimates for these joint and marginal densities. Traditional histogram-based methods are often prone to discretization errors. To overcome this limitation, we use MKDE [47] based on a normal kernel, a nonparametric method known for its accuracy in estimating the probability density function $p(x)$ of a random variable:

$$p(x) = \frac{1}{N\cdot h^d}\sum_{i=1}^{N} K\left(\frac{x-X_i}{h}\right), \tag{21}$$

where $K(\cdot)$ is the normal kernel function; the kernel's bandwidth h is optimized using a rule-of-thumb formula tailored to adapt to the dimensionality d and size N of the data:

$$h = \left(\frac{4}{d+2}\right)^{1/(d+4)}\cdot N^{-1/(d+4)}. \tag{22}$$

To promptly identify anomalous behavior in structural systems, we use a sliding window approach. Specifically, we calculate the MI between the actual measured deflections $\{X_{t-N+1}, X_{t-N}, \cdots, X_t\}$ and the predicted deflections $\{Y_{t-N+1}, Y_{t-N}, \cdots, Y_t\}$ within each window. While setting a large window size could delay the identification of anomalies, a smaller window is susceptible to noise-induced errors. Therefore, a window size of $N = 24$ hours is adopted, which strikes a balance by allowing MI to capture more information and enhancing its robustness.

However, MI inherently lacks a standardized range for effective comparison. To make MI more interpretable and to standardize its values, we utilize normalized MI (NMI) [48], which is defined as follows:

$$\lambda = \sqrt{1 - e^{-2I(X,Y)}}. \tag{23}$$

The windowed-NMI (WNMI) value to a range allows for easier comparison and interpretation, especially beneficial for deflection data with varying ranges and distributions. When a predetermined NMI threshold is set, values falling below this threshold are flagged as anomalies, facilitating real-time monitoring and timely interventions.

5. Application to a Real-World Bridge

5.1. Introduction of the SHM System on a Cable-Stayed Bridge. Situated in the urban hub of Ningbo, China, the Ningbo Bund Bridge spans an impressive 1,041 meters and features a unique architectural design. The bridge's most notable elements include a 225-meter main span and a singular, inclined triangular steel tower supported by four cable planes. The bridge's design is longitudinally asymmetrical, incorporating dual steel box girders linked by transverse beams.

Deflection in the main beam serves as a crucial indicator of the bridge's overall structural integrity, playing a pivotal role in ensuring both its longevity and operational safety. To maintain continuous surveillance of this key parameter, multiple displacement sensors have been deployed along the main span of the bridge. Operating at a sampling frequency of 1 Hz, these sensors continuously measure the relative vertical deflection values. The spatial arrangement of these measurement points, labeled PT1 through PT20, is depicted in Figure 5.

5.2. Behavior Prediction. In this study, original deflection data collected over 62 days, from April 18 to June 19, 2020, were averaged hourly to diminish the influence of transient effects. During the monitoring period, vertical wind speeds ranged from −4.72 to 4.89 m/s with an average of 0.09 m/s, and horizontal wind speeds varied from 0.00 to 12.21 m/s with an average of 1.97 m/s. The influence of such variable dynamic responses on statistically averaged deflections was considered negligible in our analysis. Moreover, as this study primarily focuses on the utilization of temporal information, only monitored deflection data and their corresponding timestamps were used, without incorporating additional inputs such as temperature data that could further enhance model performance. The patterns of temperature-induced deflections were augmented through cyclic encoding, while traffic-induced deflection patterns were enhanced through label encoding. Additionally, TPA is utilized to adjust the weights attributed to these temporal patterns, enhancing their impact on the model.

Considering that the method developed in this paper is designed for a single sensor, it ensures flexibility and can be implemented at any measurement location. Accordingly, selecting sensors near locations where potential damage is likely to occur is particularly valuable for anomaly detection. According to the literature [49], the mid-span areas of the large-span bridges studied are typically more susceptible to issues such as down-warping and box girder cracking. Therefore, three sensors located near the mid-span, specifically PT6, PT9, and PT12, have been selected as the subjects of this study. This selected dataset, visualized in Figure 6, serves a dual purpose. The data on the left side of the dashed line include the initial 1,000 hourly deflection readings, forming the training set for model calibration. The remaining data serve as the test set to evaluate the proposed predictive framework's ability to accurately capture and model the expected deflection behaviors of the bridge.

The AOA is used to fine-tune the number of hidden units in BiLSTM architectures. The hyperparameter search space is defined to range from 1 to 100 hidden units. According to the setting $n = 10$ in equation (13), the number of candidate solutions is configured as 10, indicating that AOA will evaluate 10 different configurations of hidden units at each

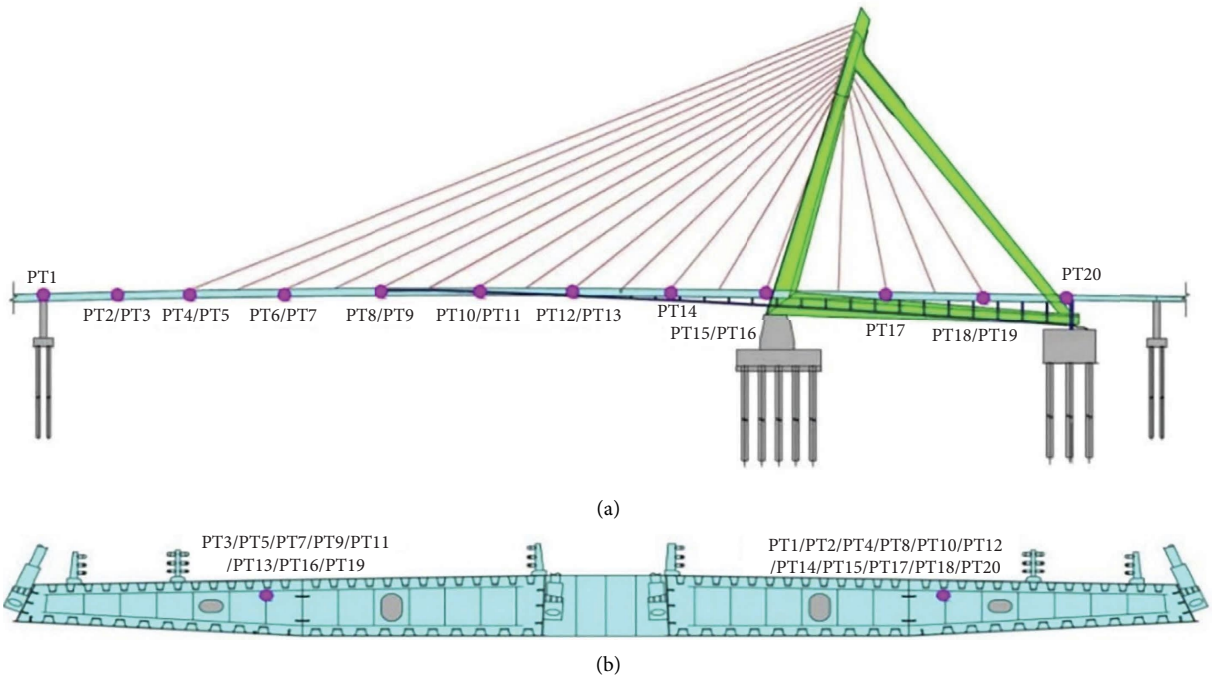

FIGURE 5: Spatial configuration of sensor arrays on the main beam: (a) schematic distribution of monitored sections; (b) detailed sensor placement within each section.

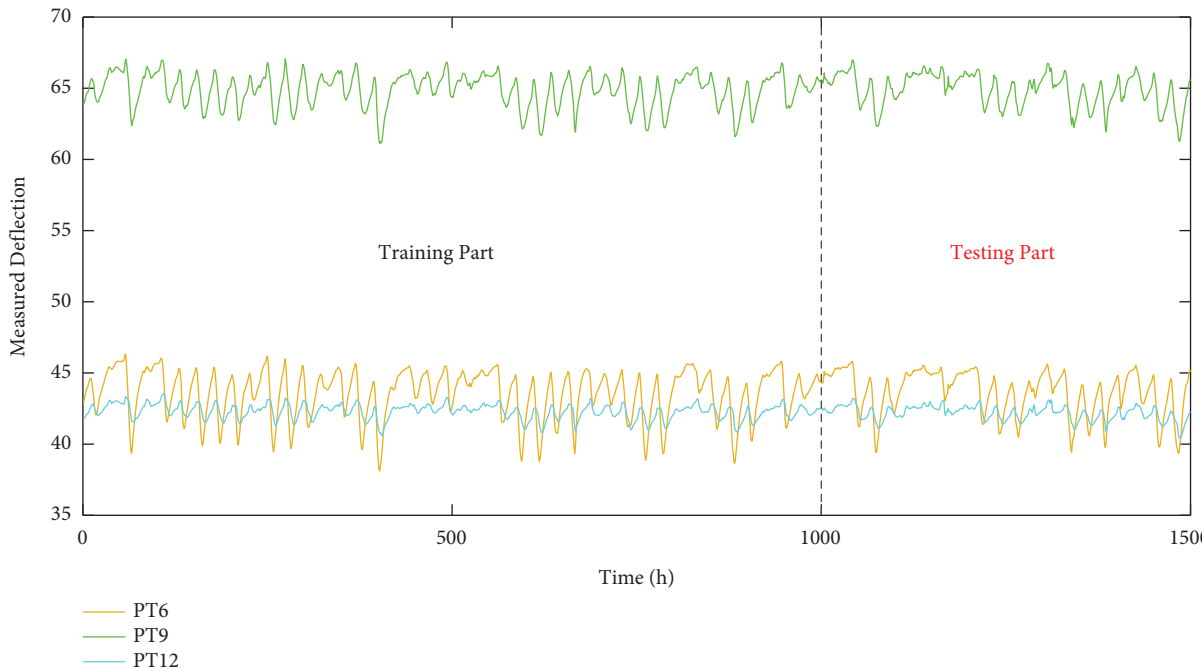

FIGURE 6: Hourly-averaged deflection measurements for selected channels: PT6, PT9, and PT12.

iterative step. The maximum number of iterations is set to 5. After 5 iterations, as demonstrated in Figure 7, optimal configurations for three distinct datasets were successfully determined, corresponding to the results at locations PT6, PT9, and PT12. The MSE was utilized as the objective function for optimization. Based on the outcomes of the optimization process, the optimal numbers of hidden units for measured deflections at locations PT6, PT9, and PT12 were identified to be 32, 45, and 18, respectively. Subsequently, using the framework with optimized parameters, the hourly-averaged deflection prediction results are presented in Figures 8–10, corresponding to the PT6, PT9, and PT12 channels, respectively. The predicted trends align well with the actual measurements, most notably around the peak

deflection values. The substantial congruence in these critical regions serves as empirical validation that the model has successfully internalized the anticipated structural behaviors.

5.3. Result Analysis. To assess the performance enhancement attributed to the encoding techniques used in this study, a control experiment was conducted with PT9 data. The performance of four model variations was compared: the proposed method, which incorporates both types of encodings; Benchmark 1, an AOA-BiLSTM-TPA model without any encoding; Benchmark 2, an AOA-BiLSTM-TPA model with only cyclic encoding; and Benchmark 3, an AOA-BiLSTM-TPA model that includes temperature data but no other encodings (temperature data inputs are depicted in Figure 11). The results, presented in Figure 12, indicate that methods integrating time encoding (the proposed method and Benchmark 2) and those adding temperature inputs align well with the measured data. In contrast, the model without any encoding (Benchmark 1) shows slightly poorer consistency. Table 1 presents the accuracy metrics averaged over 10 iterations for each method. Further analysis demonstrates that the proposed method significantly outperforms Benchmark 1 and holds a slight edge over Benchmark 2, which lacks label encoding for traffic-induced deflections. This suggests that label encoding effectively improves prediction accuracy, although the enhancement is modest. In practical applications, more detailed encoding might be needed, depending on local traffic conditions. Moreover, the proposed method and Benchmark 3, which includes additional temperature input, display varying accuracy metrics: the R^2 and RMSE are lower than those of Benchmark 3, and the MAE of the proposed method is smaller. The authors speculate that the inclusion of temperature information may enable more accurate predictions under specific conditions, such as extreme weather, accounting for the superior R^2 and RMSE performance of Benchmark 3. Nonetheless, the lower MAE suggests that the integration of time encoding with the TPA architecture has enabled the deep learning model to more effectively capture deflection evolution patterns, thus slightly reducing average errors. This indicates that cyclical temperature patterns are effectively captured through time encoding, even in the absence of direct temperature inputs, highlighting the crucial role of temporal information in enhancing predictive accuracy and compensating for the lack of direct environmental data.

To further evaluate the performance of the proposed framework, we conducted comparative analyses with several methods, including commonly used machine learning techniques such as support vector regression (SVR) [23] and Gaussian process regression (GPR) [50], as well as two state-of-the-art approaches: BiLSTM with self-attention (BiLSTM-SA) [51] and BiLSTM using standard multihead self-attention with Softmax (BiLSTM-MA) [52]. For fairness, all methods utilized AOA for hyperparameter optimization. Specifically, the hyperparameter tuning for SVR focused on the radial basis function (RBF) kernel parameters; for GPR, the length scale parameter was optimized; for BiLSTM-SA and BiLSTM-MA, the inputs are the same as the proposed method, that is, deflection data and temporal encodings, and optimization was conducted on the number of hidden units, with the number of heads in BiLSTM-MA set to 3. These comparisons were made on the same three datasets corresponding to channels PT6, PT9, and PT12, with the results illustrated in Figures 13–15. Additionally, the accuracy metrics, including RMSE, MAE, and R^2, averaged over 10 iterations for each method, are presented in Tables 2–4 for the three datasets. It is observed that in terms of RMSE and MAE, the proposed method outperformed other approaches across all three channels, indicating the smallest prediction errors. Additionally, the R^2 value is the highest among all methods, demonstrating the superior fit of the proposed framework with the actual measurements. Unlike the self-attention and multihead self-attention mechanisms that compute attention weights to select information relevant to the current timestep, the advantage of TPA lies in its ability to calculate these weights to identify important temporal patterns associated with the current timestep. Therefore, when combined with temporal encoding, TPA provides the model with more effective information conducive to learning. Notably, given the simplicity of the deflection data, the more complex multi-head attention mechanism underperformed compared to self-attention and TPA. This indicates that for simpler tasks or datasets, more complex mechanisms can lead to overfitting and reduced accuracy. It is important to note that the focus of this paper is to explore the application of temporal information in data-driven prediction of bridge deflection. Consequently, only deflection data and corresponding time encodings are used as inputs. Therefore, the model does not incorporate physical constraints through prior knowledge from physics-based modeling.

5.4. Anomaly Detection. The quantile loss function is incorporated into the proposed methodology, defining a safety margin based on the 1st and 99th percentile predictions. The predictive intervals, illustrated in Figures 16–18, correspond to the monitoring data from locations PT6, PT9, and PT12, respectively. It is noteworthy that the measurements consistently reside within the expected margins. This expected range can be utilized to detect aberrant mutations, typically in extreme conditions such as significant structural damage caused by typhoons, resulting in sudden changes in deflection. These mutations can have a lasting impact on subsequent deflections. However, such anomalous mutation values can be filtered out if they do not persist. These mutation issues are inevitable in all health monitoring systems.

Although the expected range based on quantiles aids in detecting point anomalies, it is not sensitive to slow variations in structures. Therefore, this paper uses a strategy rooted in the WNMI index to complement the expected range-based method. This dual-methodology surveillance ensures a comprehensive anomaly detection system. As shown in Figures 19–21, the persistently high values of the

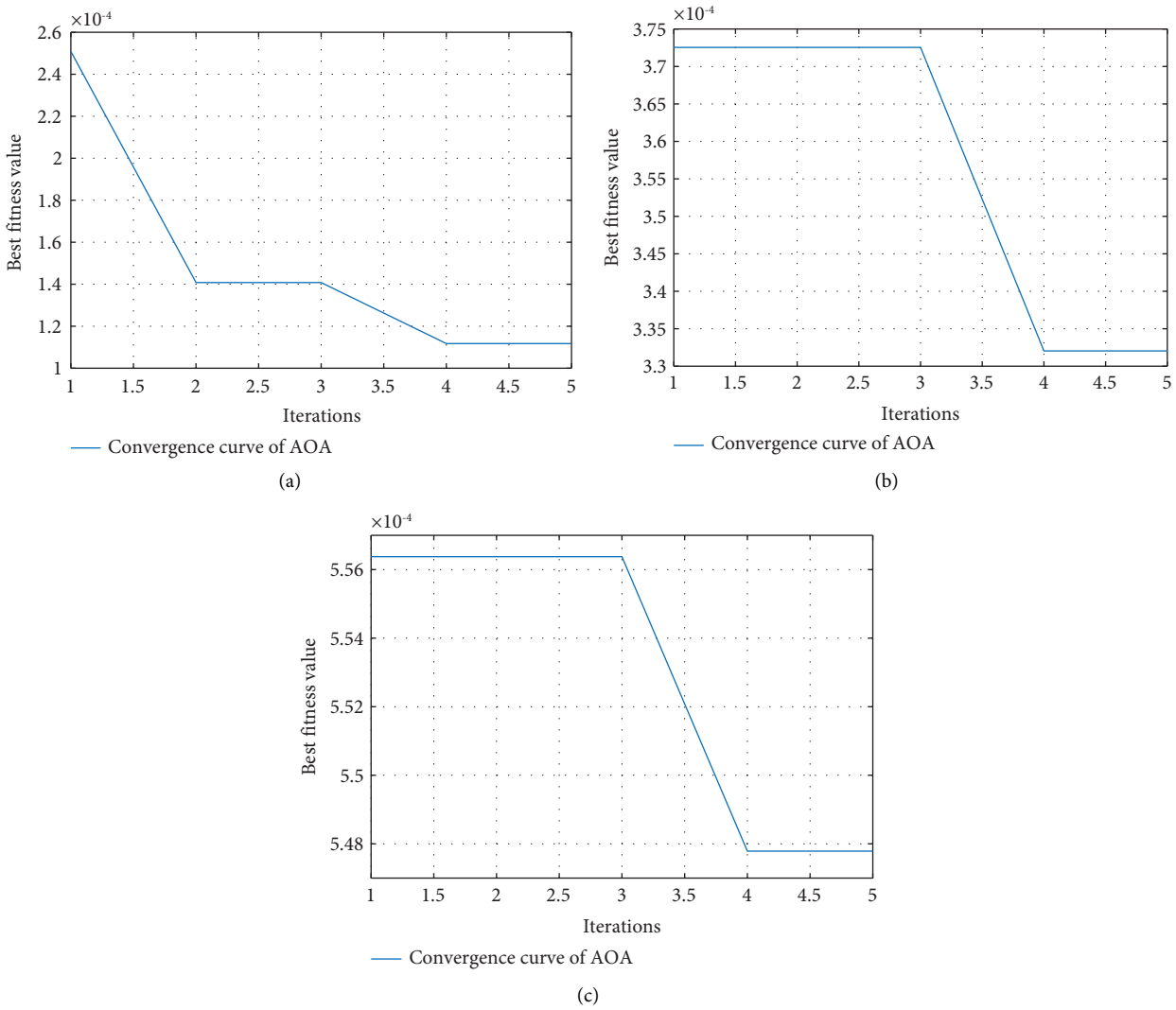

Figure 7: Determination of the optimal hyperparameter utilizing AOA for the dataset at PT6 (a), PT9 (b), and PT12 (c).

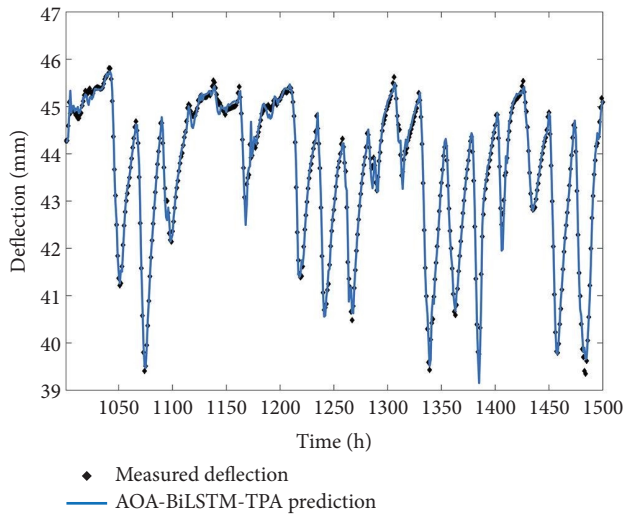

Figure 8: Predicted and measured hourly-averaged deflections at PT6.

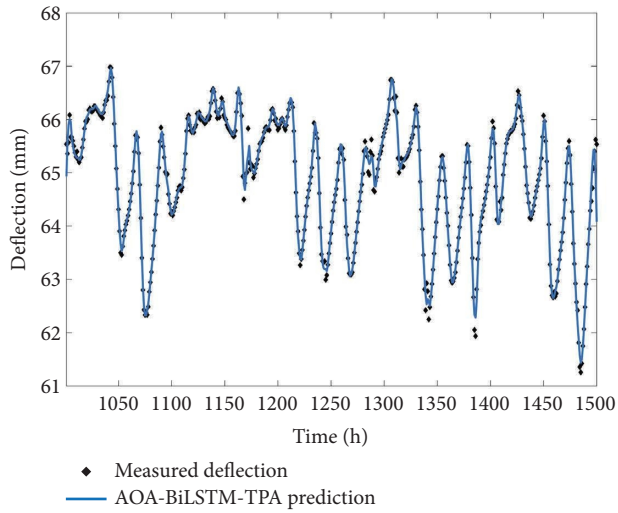

Figure 9: Predicted and measured hourly-averaged deflections at PT9.

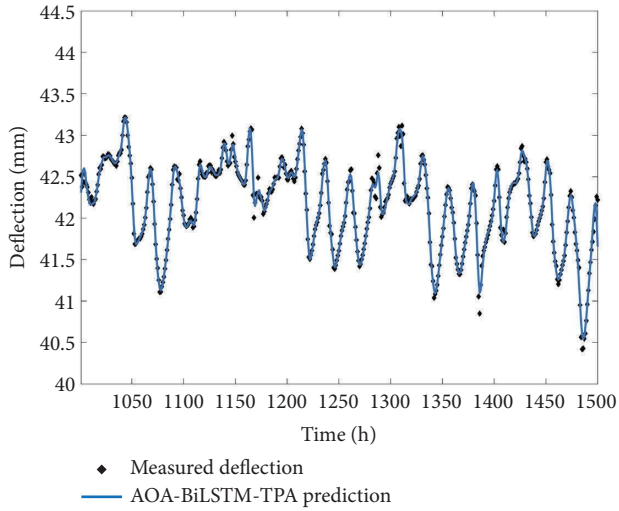

Figure 10: Predicted and measured hourly-averaged deflections at PT12.

WNMI index indicate a strong consistency between the expected and actual behavior trends of the structure. Given the observed high levels of the WNMI index (exceeding 0.9), this paper sets the WNMI threshold at 0.8 to balance sensitivity and robustness.

To further assess the anomaly detection capabilities of the proposed method, two sets of targeted simulation experiments were designed. Considering the high sensitivity of mid-span deflections to structural damages [53], which are often induced by issues such as down-warping and beam cracking typically occurring at the mid-span of structures, the PT12 located at the mid-span was selected as the subject of study. The experiments were structured as follows: (1) Extreme conditions, such as typhoon-induced bearing displacement, can cause abrupt changes in deflection, with these changes affecting subsequent deflections. This scenario serves to validate the anomaly detection method based on expected ranges. Thus, at the 1425th hour of the dataset, a gradual, abrupt increase of 2 mm was introduced, increasing by 0.2 mm per hour, to simulate an anomaly as shown in Figure 22. This increase is usually more significant. (2) Following the appearance of transverse cracks in the mid-span bottom plate, they typically develop rapidly [54], leading to a decrease in the stiffness of the beam. Additionally, the geometric deformations caused by the cracks themselves result in mid-span down-warping. This stiffness reduction results in increased deflection fluctuations. To simulate these fluctuations, Gaussian noise with a mean of 0 and a standard deviation of 0.2 mm is introduced at 1420 hours. Furthermore, to represent the progressive mid-span down-warping due to the cracks, a gradual increase of 0.01 mm per hour, starting from an initial value of 0 at 1420 hours, is implemented (see Figure 23).

The results of outlier anomaly detection based on the expected range are illustrated in Figure 22. Notably, an anomaly is marked at the 1450th hour, and the subsequent hour, 1451, is also identified as anomalous. Since the predictive model adjusts its forecasts based on observed data,

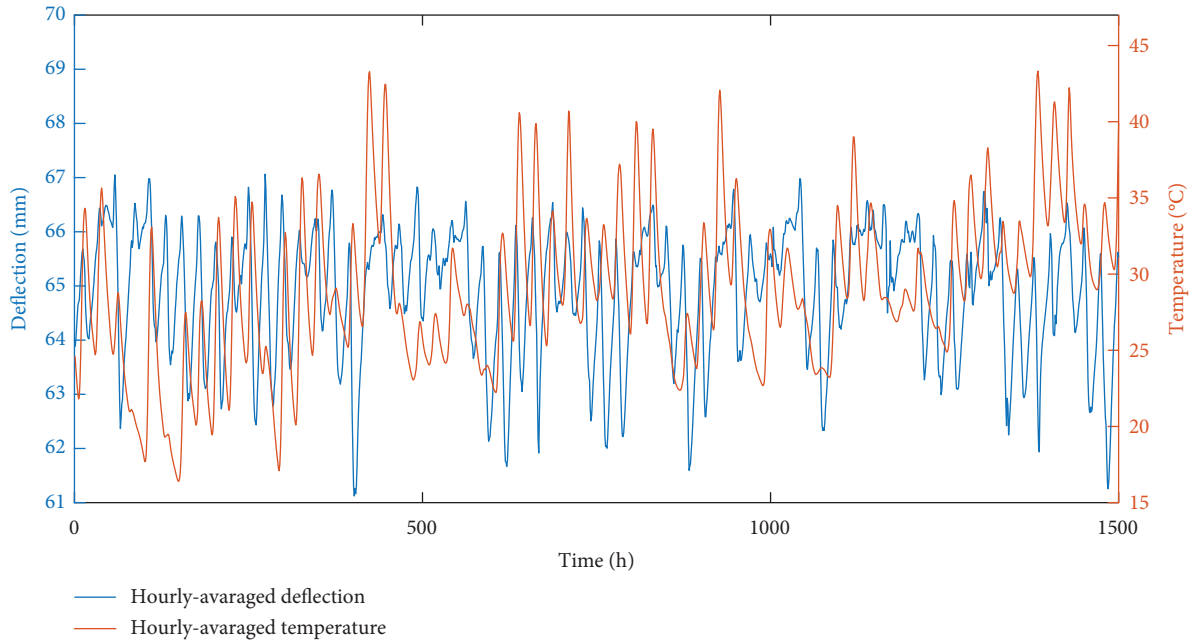

Figure 11: Hourly-averaged deflection and temperature measurements at PT9.

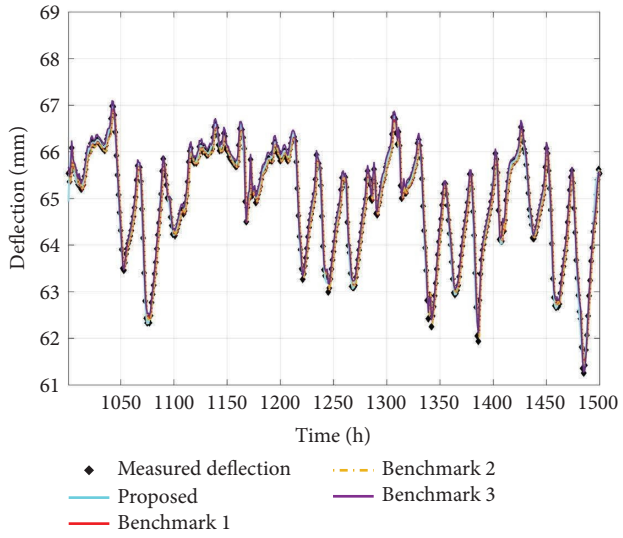

Figure 12: Comparative results of AOA-BiLSTM-TPA models with different inputs at PT9.

Table 1: Accuracy metrics of AOA-BiLSTM-TPA models with different inputs at PT9.

Metrics	Proposed	Benchmark 1	Benchmark 2	Benchmark 3
RMSE	0.1473	0.2238	0.1488	0.1329
MAE	0.1093	0.1794	0.1201	0.1155
R^2	0.9842	0.9561	0.9837	0.9855

including outlier values, and uses incremental learning for online updates, it identifies anomalies only at the moments of abrupt changes. Further analysis can be based on the average deflection level: if there are no significant changes, the structure is presumed normal; however, significant deviations suggest potential structural issues, requiring further inspection or repairs.

The prediction results for the anomalous data in simulation (2) are shown in Figure 23. Increased fluctuations following the anomaly lead to deviations between the expected structural behavior and the actual measurements. Additionally, the anomaly detection results based on the WNMI index, as seen in Figure 23, effectively identify the trend variability postanomaly, which is significantly below the threshold of 0.8. Furthermore,

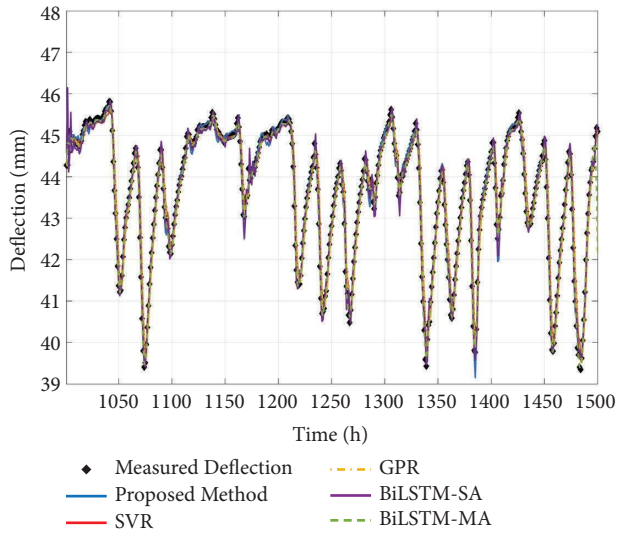

Figure 13: Comparative results of hourly-averaged deflection predictions at PT6 across different methods.

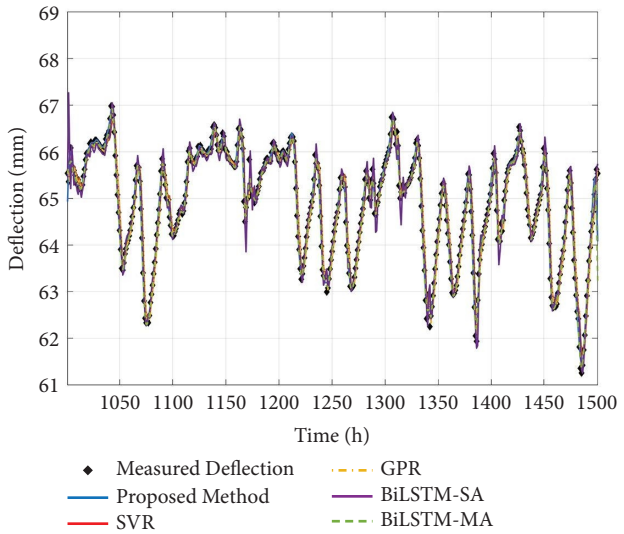

Figure 14: Comparative results of hourly-averaged deflection predictions at PT9 across different methods.

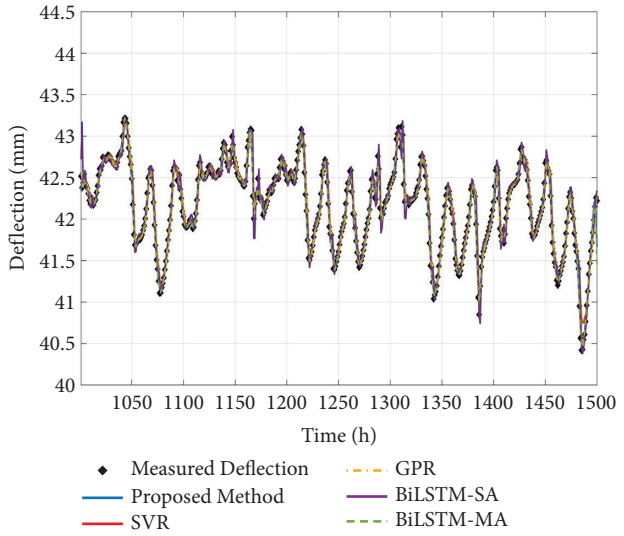

Figure 15: Comparative results of hourly-averaged deflection predictions at PT12 across different methods.

Table 2: Comparison of RMSEs, MAEs, and R^2 of deflection prediction at PT6 using different methods.

Metrics	Proposed	SVR	GPR	BiLSTM-SA	BiLSTM-MA
RMSE	**0.1550**	0.2722	0.2664	0.1721	0.1716
MAE	**0.1383**	0.1932	0.1994	0.1427	0.1541
R^2	**0.9812**	0.9438	0.9462	0.9785	0.9779

The bold values in Table 2 indicate the best performance metrics among the compared methods for each evaluation criterion.

Table 3: Comparison of RMSEs, MAEs, and R^2 of deflection prediction at PT9 using different methods.

Metrics	Proposed	SVR	GPR	BiLSTM-SA	BiLSTM-MA
RMSE	**0.1473**	0.2684	0.2479	0.1503	0.1482
MAE	**0.1093**	0.1852	0.1833	0.1260	0.1138
R^2	**0.9842**	0.9440	0.9516	0.9798	0.9819

The bold values in Table 3 indicate the best performance metrics among the compared methods for each evaluation criterion.

Table 4: Comparison of RMSEs, MAEs, and R^2 of deflection prediction at PT12 using different methods.

Metrics	Proposed	SVR	GPR	BiLSTM-SA	BiLSTM-MA
RMSE	**0.0542**	0.1349	0.1305	0.0557	0.0601
MAE	**0.0278**	0.0935	0.0927	0.0325	0.0380
R^2	**0.9878**	0.9291	0.9337	0.9827	0.9810

The bold values in Table 4 represent the best performance metrics among the compared methods for each evaluation criterion.

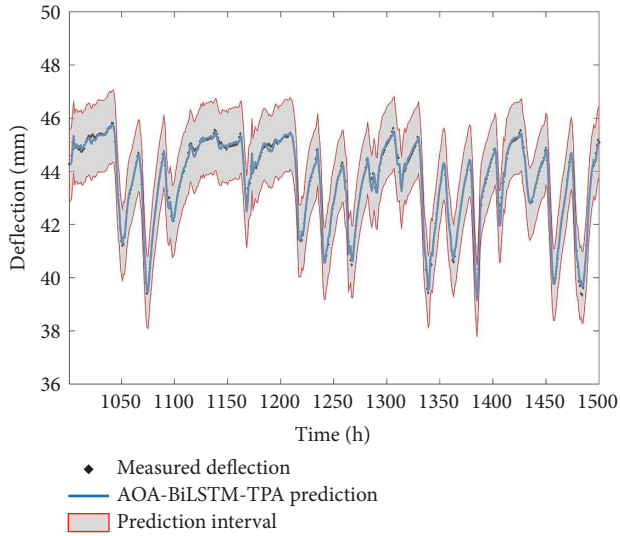

Figure 16: Prediction intervals with a safety margin based on the 1st and 99th percentile predictions at PT6.

subsequent WNMI indices display markedly denser fluctuations, differing from previous patterns. These results demonstrate that the proposed method is highly sensitive to variations in structural behavior trends, offering potential applications in detecting trend variations in deflection due to decreased stiffness.

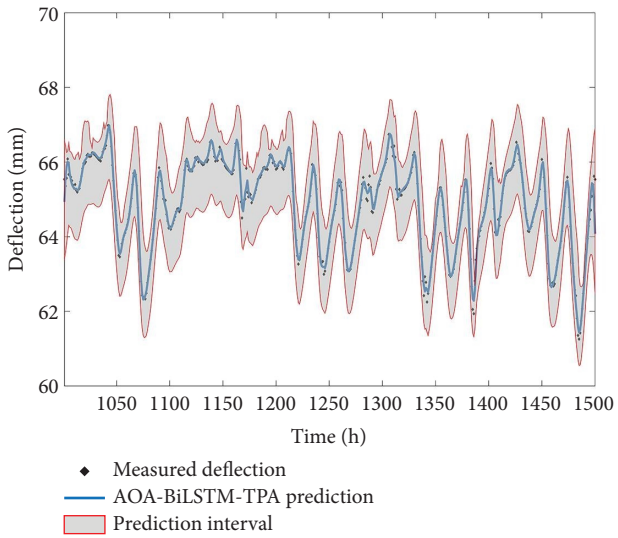

Figure 17: Prediction intervals with a safety margin based on the 1st and 99th percentile predictions at PT9.

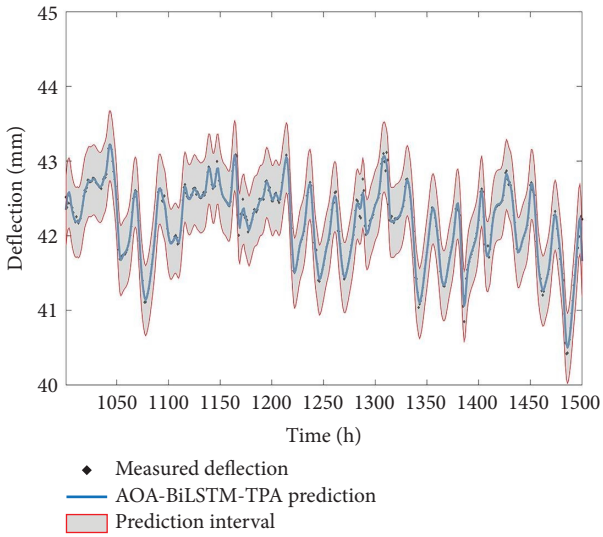

Figure 18: Prediction intervals with a safety margin based on the 1st and 99th percentile predictions at PT12.

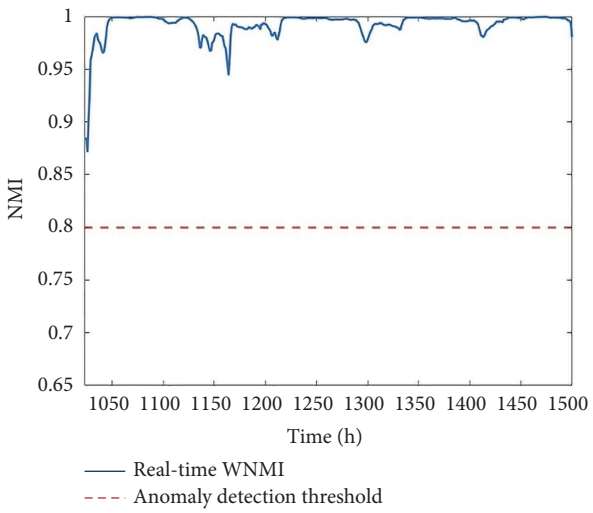

Figure 19: WNMI indices at PT6.

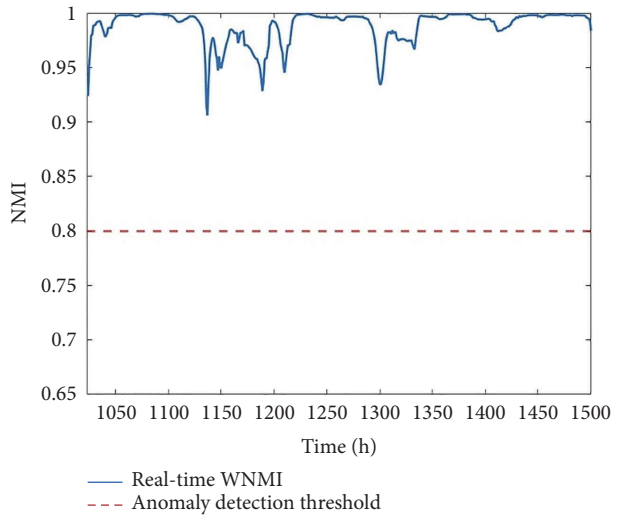

Figure 20: WNMI indices at PT9.

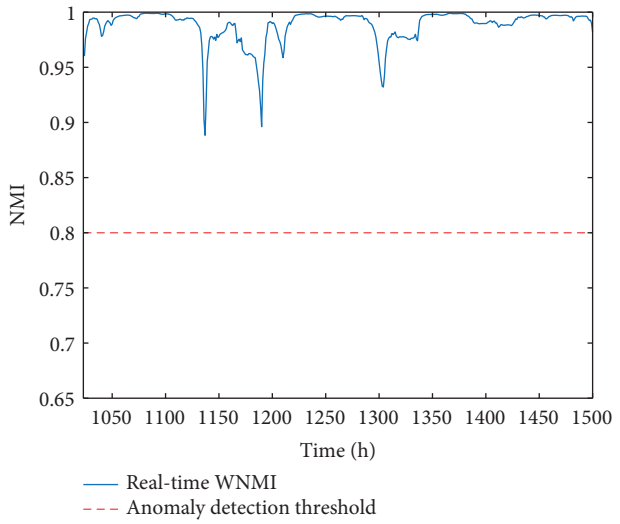

Figure 21: WNMI indices at PT12.

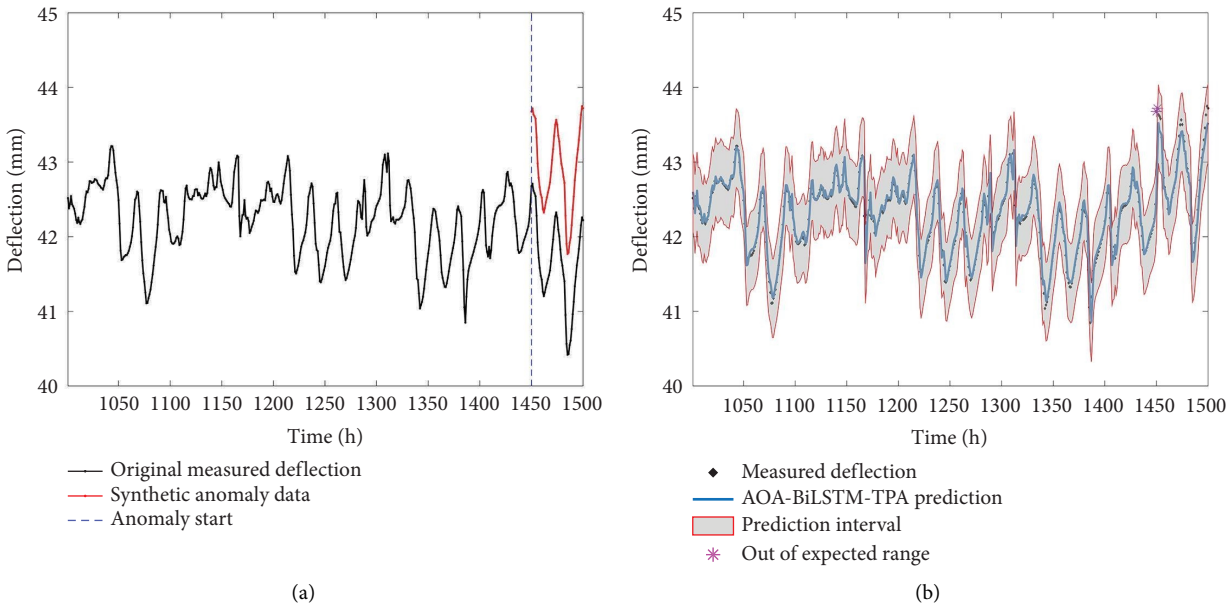

Figure 22: Simulation of abrupt anomalies under extreme conditions (a) and anomaly detection results based on the expected range (b).

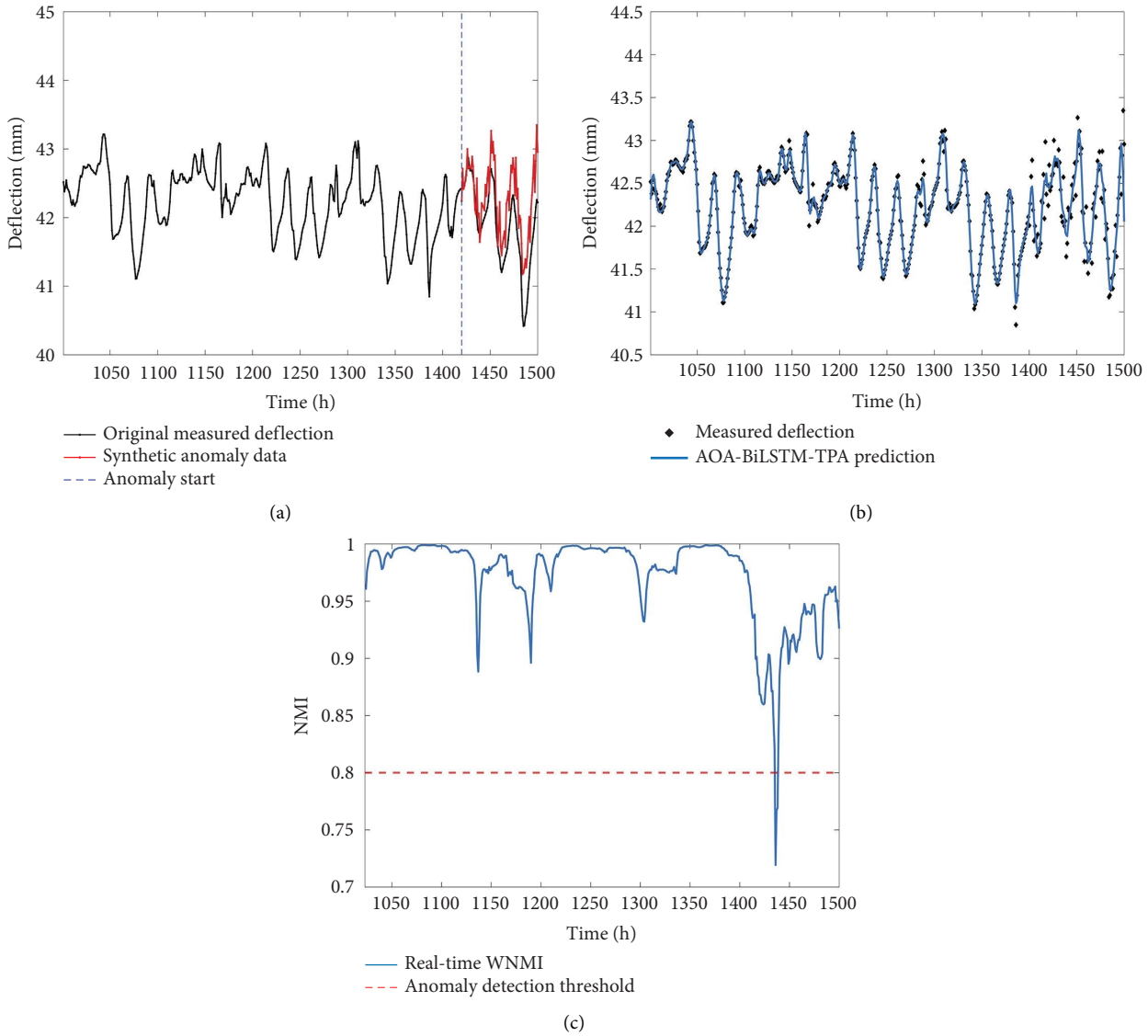

Figure 23: Simulation of increased fluctuations (due to decreased stiffness) and mid-span down-warping (a); prediction results at PT12 (b); WNMI indices for anomaly detection at PT12 (c).

6. Conclusions

This study presents an innovative framework for real-time anomaly detection in operational bridge deflections. Using cyclic encoding, 24-hour timestamps are represented as points on a circle to enhance the capture of temperature-induced deflection behaviors. Label encoding is applied to encode morning rush hours (Beijing time 07:00–09:00) and evening rush hours (Beijing time 15:00–19:00) to improve learning of traffic-induced deflection patterns. These temporal patterns are learned through the integration of a BiLSTM model with the TPA mechanism. Based on this, an incremental learning strategy using experience replay is applied, allowing the model to update in real time. The selection of hyperparameters is enhanced by the AOA method, improving the objectivity and accuracy of these parameters. Another innovative aspect is the dual-strategy approach to anomaly detection, which increases robustness: the quantile loss function provides an expected range for deflection to identify isolated irregularities, and trend variability is detected using the MKDE-based WNMI.

The efficacy of the proposed model was validated using measurement data of deflections from three channels of an operational cable-stayed bridge. The model demonstrated superior performance in capturing deflection patterns that vary with time. When compared to several alternative methods, including SVR, GPR, BiLSTM with self-attention, and BiLSTM with multihead attention, the proposed model yielded the highest accuracy across all three datasets. The determination coefficients (R^2) for all three sets exceeded 0.98, indicating excellent agreement between the predicted and actual deflection data. The efficacy of the proposed anomaly detection method was further validated through two simulation experiments. The first simulation tested the

method's ability to detect sudden deflection changes under extreme conditions and accurately identified anomalies using expected range-based analysis. The second simulation assessed deflection trend variability caused by decreased structural stiffness, with anomalies detected using the WNMI index.

These findings highlight the critical role of the time dimension in providing key insights for anomaly detection in SHM. To emphasize this, the current study used only monitored deflection and its corresponding temporal data as inputs. Future research will expand the model's robustness by incorporating additional physical parameters. Additionally, there is a need to integrate physical constraints with data-driven methods through physics-based modeling, such as incorporating the temperature-displacement relationship (TDR) into the model, to enhance its accuracy and interpretability.

Data Availability

The data used to support the findings of this study are available upon request from the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors wish to acknowledge the financial support provided by the National Natural Science Foundation of China (52278313 and 52411540031) and the Technology Cooperation Project of Shanghai Qizhi Institute Cooperation (SQZ202310).

References

[1] D. K. Devendiran and S. Banerjee, "Influence of combined corrosion–fatigue deterioration on life-cycle resilience of RC bridges," *Journal of Bridge Engineering*, vol. 28, no. 5, Article ID 04023014, 2023.

[2] I. Bondarenko, V. Lukoševičius, R. Keršys, and L. Neduzha, "Innovative trends in railway condition monitoring," *Transportation Research Procedia*, vol. 77, pp. 10–17, 2024.

[3] D. Diamantidis, P. Tanner, M. Holicky, H. O. Madsen, and M. Sykora, "On reliability assessment of existing structures," *Structural Safety*, Article ID 102452, 2024.

[4] A. Jiménez Rios, V. Plevris, and M. Nogal, "Bridge management through digital twin-based anomaly detection systems: a systematic review," *Frontiers in Built Environment*, vol. 9, p. 61, 2023.

[5] L. Sun, Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah, "Review of bridge structural health monitoring aided by big data and artificial intelligence: from condition assessment to damage detection," *Journal of Structural Engineering*, vol. 146, no. 5, Article ID 04020073, 2020.

[6] G. Qu, L. Sun, and H. Huang, "Bridge performance prediction based on a novel SHM-data assimilation approach considering cyclicity," *Structural Control and Health Monitoring*, vol. 2023, pp. 1–16, 2023.

[7] M. Jasiński, P. Łaziński, and D. Piotrowski, "The concept of creating digital twins of bridges using load tests," *Sensors*, vol. 23, no. 17, p. 7349, 2023.

[8] G. E. Vazquez B, J. R. Gaxiola-Camacho, R. Bennett, G. M. Guzman-Acevedo, and I. E. Gaxiola-Camacho, "Structural evaluation of dynamic and semi-static displacements of the Juarez Bridge using GPS technology," *Measurement*, vol. 110, pp. 146–153, 2017.

[9] G. Qu and L. Sun, "Performance prediction for steel bridges using SHM data and bayesian dynamic regression linear model: a novel approach," *Journal of Bridge Engineering*, vol. 29, no. 7, Article ID 04024044, 2024.

[10] C. Adam, R. Heuer, and F. Ziegler, "Reliable dynamic analysis of an uncertain compound bridge under traffic loads," *Acta Mechanica*, vol. 223, no. 8, pp. 1567–1581, 2012.

[11] N. Subramanian, "I-35W Mississippi river bridge failure–Is it a wake up call?" *Indian Concrete Journal*, vol. 19, no. 9, pp. 29–38, 2008.

[12] M. Z. Naser, "Can past failures help identify vulnerable bridges to extreme events? A biomimetical machine learning approach," *Engineering with Computers*, vol. 37, no. 2, pp. 1099–1131, 2021.

[13] Z. Sun, J. Santos, and E. Caetano, "Data driven prediction and interpretation of fatigue damage in a road rail suspension bridge considering multiple loads," *Structural Control and Health Monitoring*, vol. 29, no. 9, Article ID e2997, 2022.

[14] M. M. Futai, T. N. Bittencourt, R. R. Santos et al., "Utilization of digital twins for bridge inspection, monitoring and maintenance," *Lecture Notes in Civil Engineering*, vol. 1, pp. 166–173, 2022.

[15] D. La Mazza, F. Basone, M. Longo, P. Darò, and A. Cigada, "Anomaly detection through long-term SHM: some interesting cases on bridges," in *Proceedings of the Dynamics of Civil Structures, Volume 2: Proceedings of the 40th IMAC, A Conference and Exposition on Structural Dynamics*, Berlin, Germany, June 2022.

[16] D. Kim and J. Lee, "A review of physics informed neural networks for multiscale analysis and inverse problems," *Multiscale Science and Engineering*, vol. 52, 2024.

[17] D. Nath, D. R. Neog, D. R. Neog, and S. S. Gautam, "Application of machine learning and deep learning in finite element analysis: a comprehensive review," *Archives of Computational Methods in Engineering*, vol. 12, pp. 1–40, 2024.

[18] L. H. Nguyen and J. A. Goulet, "Real time anomaly detection with Bayesian dynamic linear models," *Structural Control and Health Monitoring*, vol. 26, no. 9, p. e2404, 2019.

[19] Q. A. Wang, Y. Dai, Z. G. Ma et al., "Towards probabilistic data driven damage detection in SHM using sparse Bayesian learning scheme," *Structural Control and Health Monitoring*, vol. 29, no. 11, Article ID e3070, 2022.

[20] Z. Yue, Y. Ding, H. Zhao, and Z. Wang, "Mechanics-Guided optimization of an LSTM network for Real-Time modeling of Temperature-Induced deflection of a Cable-Stayed bridge," *Engineering Structures*, vol. 252, Article ID 113619, 2022.

[21] N. Manzini, A. Orcesi, C. Thom et al., "Machine learning models applied to a GNSS sensor network for automated bridge anomaly detection," *Journal of Structural Engineering*, vol. 148, no. 11, Article ID 04022171, 2022.

[22] H. Zhang, J. Lin, J. Hua, F. Gao, and T. Tong, "Data anomaly detection for bridge SHM based on CNN combined with statistic features," *Journal of Nondestructive Evaluation*, vol. 41, no. 1, p. 28, 2022.

[23] R. Kromanis and P. Kripakaran, "Support vector regression for anomaly detection from measurement histories," *Advanced Engineering Informatics*, vol. 27, no. 4, pp. 486–495, 2013.

[24] J. Xi Yang, F. Yang, L. Zhang et al., "Bridge health anomaly detection using deep support vector data description," *Neurocomputing*, vol. 444, pp. 170–178, 2021.

[25] A. Dalvi, "Performance of one-class Support Vector Machine (SVM) in detection of anomalies in the bridge data," Springer, Berlin, Germany, 2017, PhD diss.

[26] P. Miao, G. Xing, S. Ma, and T. Srimahachota, "Deep learning–based inspection data mining and derived information fusion for enhanced bridge deterioration assessment," *Journal of Bridge Engineering*, vol. 28, no. 8, Article ID 04023048, 2023.

[27] M. Arul and A. Kareem, "Data anomaly detection for structural health monitoring of bridges using shapelet transform," 2020.

[28] Z. Yue, Y. Ding, and H. Zhao, "Deep learning-based minute-scale digital prediction model of temperature-induced deflection of a cable-stayed bridge: case study," *Journal of Bridge Engineering*, vol. 26, no. 6, Article ID 05021004, 2021.

[29] Y. Zhu, D. Sun, and M. Shuang, "Investigation of temperature-induced effect on rail-road suspension bridges during operation," *Journal of Constructional Steel Research*, vol. 215, Article ID 108542, 2024.

[30] G. W. Roberts, C. Brown, and X. Meng, "Deflection monitoring of the forth road bridge by GPS," in *Proceedings of the 18th International Technical Meeting of the Satellite*, Honolulu, Hawaii, September, 2005.

[31] X. Zheng, T. H. Yi, J. W. Zhong, and D. H. Yang, "Rapid evaluation of load-carrying capacity of long-span bridges using limited testing vehicles," *Journal of Bridge Engineering*, vol. 27, no. 4, Article ID 04022008, 2022.

[32] S. Jia, M. Akiyama, B. Han, and D. M. Frangopol, "Probabilistic structural identification and condition assessment of prestressed concrete bridges based on Bayesian inference using deflection measurements," *Structure and Infrastructure Engineering*, vol. 20, pp. 131–147, 2023.

[33] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Computation*, vol. 31, no. 7, pp. 1235–1270, 2019.

[34] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM--a tutorial into long short-term memory recurrent neural networks," 2019.

[35] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "The performance of LSTM and BiLSTM in forecasting time series," in *Proceedings of the 2019 IEEE International conference on big data (Big Data)*, pp. 3285–3292, Los Angeles, CA, USA, December, 2019.

[36] S. Y. Shih, F. K. Sun, and H. Y. Lee, "Temporal pattern attention for multivariate time series forecasting," *Machine Learning*, vol. 108, no. 8-9, pp. 1421–1441, 2019.

[37] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, and A. H. Gandomi, "The arithmetic optimization algorithm," *Computer Methods in Applied Mechanics and Engineering*, vol. 376, Article ID 113609, 2021.

[38] L. R. Medsker and L. C. Jain, "Recurrent neural networks," *Design and Applications*, vol. 5, no. 64-67, p. 2, 2001.

[39] J. H. Holland, "Genetic algorithms," *Scientific American*, vol. 267, no. 1, pp. 66–72, 1992.

[40] J. Kennedy and R. Eberhart, "Particle swarm optimization," *Proceedings of ICNN'95-international conference on neural networks*, vol. 4, pp. 1942–1948, 1995.

[41] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," 2014.

[42] G. Petit, M. Soumm, E. Feillet et al., "An analysis of initial training strategies for exemplar-free class-incremental learning," in *Proceedings of the 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1837–1847, Waikoloa, HI, USA, June 2024.

[43] J. Shen, T. Ma, D. Song, and F. Xu, "Incremental learning BiLSTM based on dynamic proportional adjustment mechanism and experience replay for quantitative detection of blade crack propagation," *Structural Health Monitoring*, vol. 23, no. 2, pp. 733–749, 2024.

[44] V. K. Patidar, R. Wadhvani, S. Shukla, M. Gupta, and M. Gyanchandani, "Quantile regression comprehensive in machine learning: a review," in *Proceedings of the 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp. 1–6, Bhopal, India, February, 2023.

[45] Q. Ren, M. Li, and Y. Shen, "A new interval prediction method for displacement behavior of concrete dams based on gradient boosted quantile regression," *Structural Control and Health Monitoring*, vol. 29, no. 1, Article ID e2859, 2022.

[46] J. R. Vergara and P. A. Estévez, "A review of feature selection methods based on mutual information," *Neural Computing and Applications*, vol. 24, no. 1, pp. 175–186, 2014.

[47] L. Batina, B. Gierlichs, E. Prouff, M. Rivain, F. X. Standaert, and N. Veyrat-Charvillon, "Mutual information analysis: a comprehensive study," *Journal of Cryptology*, vol. 24, no. 2, pp. 269–291, 2011.

[48] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Transactions on Neural Networks*, vol. 20, no. 2, pp. 189–201, 2009.

[49] G. Xu, X. Yu, J. Li, Z. Chen, and X. Zong, "Research on long-term deformation monitoring of PC box girder bridge strengthened by a cable-stayed system," *IOP Conference Series: Earth and Environmental Science*, vol. 769, no. 3, p. 032028, Article ID 032028, 2021.

[50] F. Ghodoosi, A. Bagchi, T. Zayed, and M. R. Hosseini, "Method for developing and updating deterioration models for concrete bridge decks using GPR data," *Automation in Construction*, vol. 91, pp. 133–141, 2018.

[51] J. Liu, X. Tang, and X. Guan, "Grain protein function prediction based on self-attention mechanism and bidirectional LSTM," *Briefings in Bioinformatics*, vol. 24, no. 1, 2023.

[52] E. Voita, D. Talbot, F. Moiseev, R. Sennrich, and I. Titov, "Analyzing multi-head self-attention: specialized heads do the heavy lifting, the rest can be pruned," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 2019.

[53] H. W. Shih, D. P. Thambiratnam, and T. H. T. Chan, "Vibration based structural damage detection in flexural members using multi-criteria approach," *Journal of Sound and Vibration*, vol. 323, no. 3-5, pp. 645–661, 2009.

[54] B. Wan, C. Foley, and J. Komp, *Concrete Cracking in New Bridge Decks and Overlays*, Marquette University, Milwaukee, WI, USA, 2010.