

Digital Model of Temperature-induced Deflection of Bridge Driven via Deep Learning

* Junxiao Guo ¹⁾ and Youliang Ding ²⁾

^{1), 2)} Southeast University, Nanjing 210-096, China

¹⁾ 220211291@seu.edu.cn

ABSTRACT

Based on big data, we can build a regression model between a temperature field and a temperature-induced deflection to provide a control group representing the service performance of bridges, which has a positive effect on the full life cycle maintenance of bridges. However, the spatial temperature information of a cable-stayed bridge is difficult to describe. To establish a regression model with high precision, the improved PCA-LGBM (principal component analysis and light gradient boosting machine) algorithm is proposed to extract the main temperature features that can reflect the spatial temperature information as accurately and efficiently as possible. Then, in this article, we searched for a suitable digital tool for modeling the regressive relationship between the temperature variables and the temperature-induced deflection of a cable-stayed bridge. The precision of the backpropagation neural network (BPNN) model has been improved, but it is still unsatisfactory. The nested long short-term memory (NLSTM) model improves the nonlinear expression ability of the regression model and is more precise than BPNN models and the classical LSTM. The architecture of the NLSTM network is optimized for high precision and to avoid the waste of computational costs. Based on the four main temperature features extracted via the PCA-LGBM, the NLSTM network with double hidden layers and 256 hidden units in each hidden layer has much higher precision than the other regression models. For the NLSTM regression model of the temperature-induced deflection of a cable-stayed bridge, the mean absolute error is only 4.76 mm, and the mean square error is only 18.57 mm². The control value of the NLSTM regression model is precise and thus provides the potential for early detection of bridge anomalies. This article can provide reference processes and a data extraction algorithm for deflection modeling of other cable-stayed bridges.

1. INTRODUCTION

Cable-stayed bridges are common long-span bridges that are increasingly becoming widely used. To ensure the safety of cable stayed bridges for the full life cycle,

¹⁾ Graduate Student

²⁾ Professor

structural health monitoring (SHM) systems can be installed to monitor the use status of bridges to avert economic and safety risks caused by destructive cable-stayed bridge accidents. (Sun et al. 2020; Miyamoto & Motoshita 2017) In the full life cycle of the structure, de-deflection is a control index reflecting mechanical properties. (Xia 2014) In the operation phase, the most critical factors affecting the vertical deflection of the main girder are temperature and vehicle load; however, the impact of thermal load is more significant. (Zhao et al. 2019; Ding et al. 2013) Accordingly, we can establish a correlation model between temperature and deflection induced by temperature. By inputting the real-time temperature data from the SHM system, we can obtain the output data of deflection induced by temperature and use the output data as control values. The normal working state of the bridge can then be evaluated according to the difference between its deflection and the control value. Efficient evaluation of bridges will facilitate the prolonging life of bridges as early as possible, and the longer bridge service life will obviously benefit the sustainability of infrastructures. Therefore, a high-precision model that provides reliable deflection control values is urgently required. (Xu et al. 2021)

Naturally, the temperature of the girder affects its deflection, which is also affected by the temperature of the bridge tower. (Zhou et al. 2020; Yi et al. 2020) In the existing analytic model, even if the composite effect of multiple temperatures is considered, large errors occur. (Zhou & Sun. 2019) Benefiting from the wide application of SHM systems, a large amount of environmental information and bridge response data were collected. (Wang et al. 2020) Therefore, the data obtained via the SHM system is obviously big data that contain potential information about the service performance of the bridge. (Wang et al. 2020; Tu et al. 2017)

The information contained in bridge monitoring data is very valuable; therefore, utilizing big data to promote bridge engineering research has become the general trend. (Cai & Mahadevan. 2018) To ensure that the mapping relationship between temperature and temperature-induced deflection can be precisely expressed through regression modeling, deep learning technology should be applied to satisfy the high-precision regression model.

Firstly, for optimal precision, the temperature feature extraction method combining two intelligent algorithms is proposed in this paper. The principal component analysis (PCA) algorithm is usually used for restructuring temperature field data into several main temperature features; the restructured data features obtained via PCA can better adapt to nonlinear modeling, but the redundant features must be re-moved. (Huang et al. 2016) Therefore, in this paper, light gradient boosting machine (LGBM), which belongs to machine learning technology and is able to better explain the nonlinear connections between data features, is further used to select the most useful features from the data information reconstructed via PCA. (Punmiya & choe 2019)

Then, to determine the appropriate deep learning modeling tool, this paper tested different neural networks and the network with the best performance was selected. The traditional backpropagation neural network (BPNN) with the typical back propagation characteristic was tested. (Wang et al. 2015; Singh et al. 2022) The advanced nested long short-term memory (NLSTM) network that can express highly nonlinear temporal relationship was tested. (Ma et al. 2021) And the classic LSTM network was also tested to prove that NLSTM is more suitable for the mapping model in this paper. (Smagulova

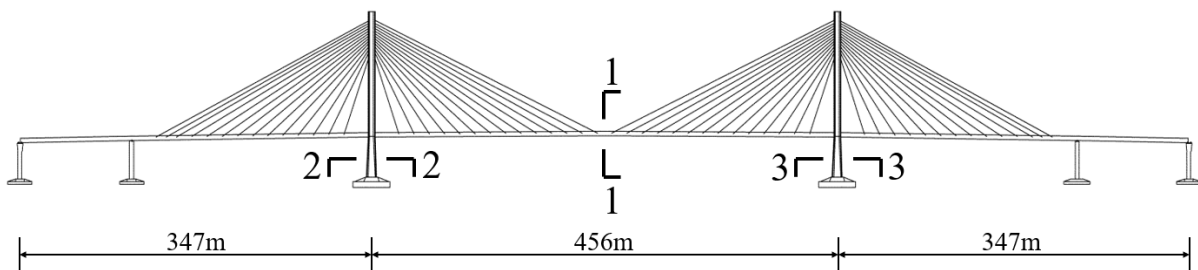
& James 2019) Meanwhile, the paper explores the optimal architecture of neural networks with the comprehensive consideration of effect and cost.

Finally, the digital regression model built by using the deep neural network can achieve a very high precision, thus providing a valuable control group for bridge deflection; therefore, the abnormal state of the bridge can be detected as early as possible, which has a positive significance for bridge health monitoring.

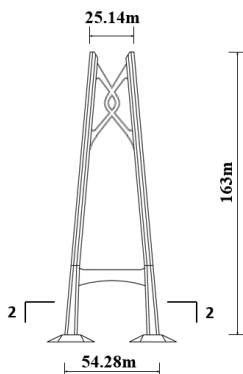
2. EXTRACTION OF TEMPERATURE FEATURES AND TEMPERATURE-INDUCED DEFLECTION

2.1 Bridge and Monitoring System

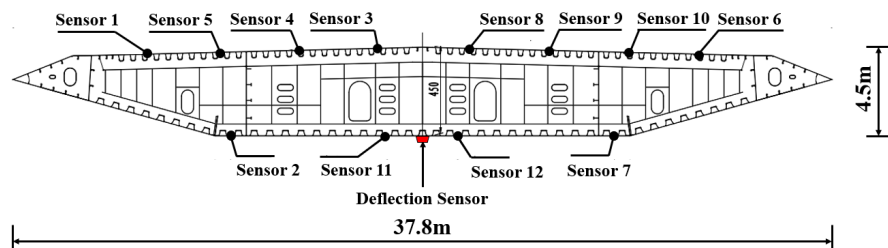
This research is based on data from the SHM system of a sea-crossing bridge between Macau and Hong Kong. As shown in Fig. 1(a), the bridge has a total length of 1150 m and a main span of 456 m. The main girders are steel box beams measuring 36 m in width and 4.6 m in height. As shown in Fig. 1(b), the bridge towers adopt the concrete structure with a box section, and the two towers are 163 m high. As shown in Fig. 1(c), there are twelve temperature sensors installed in the main girder and these temperature sensors are denoted as sensor 1 to sensor 12, and there is a deflection sensor in the center of the main girder. As shown in Fig. 1(d) and Fig. 1(e), there are eight temperature sensors installed in the concrete structure of the two bridge towers, and these temperature sensors are denoted as sensor 13 to sensor 20.



(a) Elevation of the cable-stayed bridge



(b) Bridge tower



(c) Section 1-1

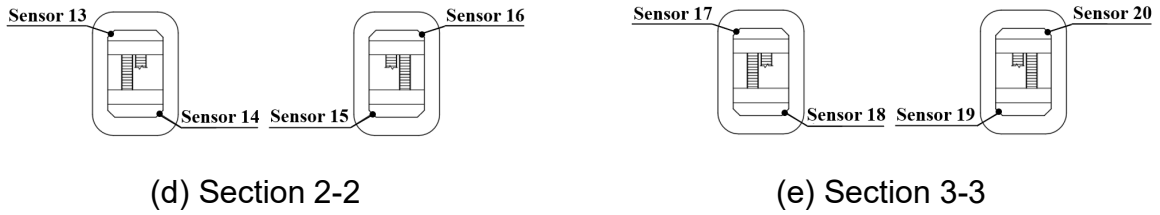


Fig. 1 Positions of the temperature sensors installed on the bridge

2.2 Temperature Information

Fig. 2 shows the temperature data of sensors 4, 7, 11, and 18 during a one-year period after the ten-minute average processing. The temperature variables of the 20 sensors after processing are named $T_1 \sim T_{20}$. Taking all the temperature variables as the input data for modeling not only wastes a lot of computing cost but may have a negative impact because there exists the obvious information redundancy between the temperature measured via different sensors. The datasets collected via these temperature sensors are similar but have independent significance. Therefore, these data cannot be randomly selected or discarded; thus, all the temperature data need to be restructured in order to extract the main temperature features as the input data of the regression model.

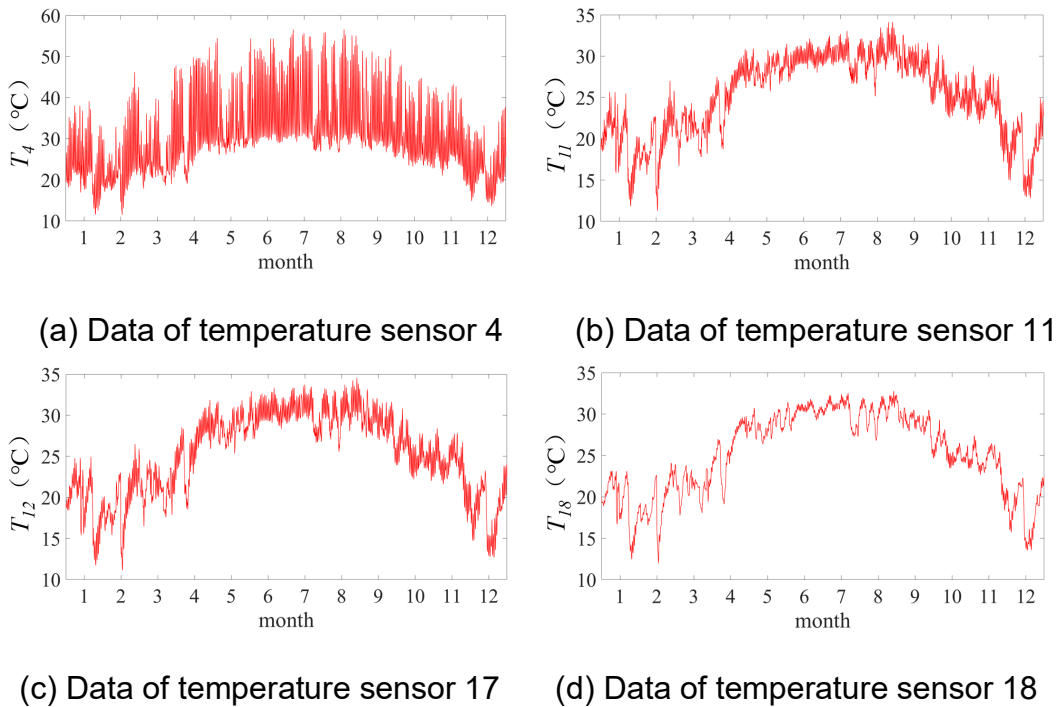


Fig. 2 Time-history curves of the temperature measured via different sensors

2.3 Extraction of Temperature Features

For the extraction of temperature features, the Principal Component Analysis algorithm is mainly combined with Light Gradient Boosting Machine algorithm to reconstruct and screen the temperature features.

Information Reconstruction via Principal Component Analysis Principal Component Analysis (PCA) can recombine the multiple measured information into new variables that are independent of each other. (Erichson et al. 2018) Therefore, it can reconstruct the data according to the actual situation. Some of the reconstructed variables will contain a lot of effective information, while the other variables are redundant information that can be removed.

If there are n parameters, the centralized parameters can form a matrix $W = [w_1, w_2, w_3, \dots, w_n]$. The w_n is described by Eq. (1):

$$w_n = [w_{n1}, w_{n2}, w_{n3}, \dots, w_{nk}]^T \quad (1)$$

where k is the amount of data.

The orthogonal group that forms a new mapping space is $\{u_1, u_2, u_3, \dots, u_k\}$, and the projection of the original characteristic parameter w_i in u_j can be described as $w_i^T u_j$. The variance in the original data after the projection in u_j can be described by Eq. (2):

$$J_j = \frac{1}{n} \sum_{i=1}^n (w_i^T \cdot u_j)^2 = \frac{u_j^T}{n} W \cdot W^T u_j \quad (2)$$

S is the covariance matrix of W . It is described by Eq. (3):

$$S = W \cdot W^T / n \quad (3)$$

The solution to the projection space can be converted into Eq. (4):

$$\begin{cases} \max J_j = u_j^T S u_j \\ \text{s.t. } u_j^T u_j = 1 \end{cases} \quad (4)$$

According to the Lagrangri multiplier method, Eq. (5) can be constructed:

$$F(u_j) = u_j^T S u_j + \lambda_j (1 - u_j^T u_j) \quad (5)$$

Taking the derivative of function F , the result is described by Eq. (6):

$$\frac{\partial F}{\partial u_j} = 2S u_j - 2\lambda_j u_j = 0 \Rightarrow S u_j = \lambda_j u_j \quad (6)$$

where λ_j is the eigenvalue and u_j is the eigenvector of S .

The eigenvectors are composed of a matrix $U = [u_1, u_2, u_3, \dots, u_j]$. The matrix of reconstructed information can be described by Eq. (7):

$$B = [b_1, b_2, b_3, \dots, b_j] = W * U \quad (7)$$

The column vector \mathbf{b}_j is the reconstructed vector.

Based on the PCA algorithm, the principal components are orthogonal and non-correlated. Through this method, the reconstruction of the temperature variables can be performed efficiently and accurately. The temperature variables obtained by the reconstruction are named $T_1' \sim T_{20}'$. In the regression modeling of temperature deflection, these temperature variables reconstructed via PCA have different information abundances; therefore, the relevant algorithm must be further used to analyze the information contribution of different features in order to retain the most useful information and remove redundant data. Subsequent work revolves around the light gradient boosting machine framework.

Analysis of the Reconstructed Information via Light Gradient Boosting Machine GBDT (Gradient Boosting Decision Tree) is the classical and reliable analytical method belonging to machine learning. As the improved framework based on the GBDT algorithm, LGBM can reduce the computational cost by parallel training and thus be widely used in data mining. (Ju et al. 2019)

Similar to the GBDT, LGBM (Light Gradient Boosting Machine) implements the decision tree by dividing the histogram of the original data. However, LGBM has been further optimized by replacing the traditional level-wise algorithm by the leaf-wise algorithm with depth constraints. Fig. 3 shows the difference between the level-wise algorithm and the leaf-wise algorithm. Therefore, the LGBM algorithm can accurately sort the importance of different variables by the decision tree scheme.

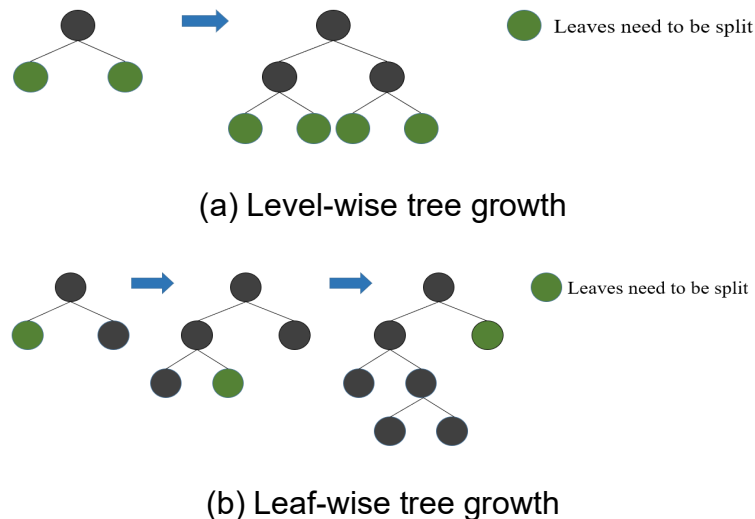


Fig. 3 The difference between level-wise decision trees and leaf-wise decision trees

The reconstructed temperature variables obtained via PCA are completely uncorrelated. However, the information contribution of the different variables to the deflection regression is not the same. As the number of input features to the regression model increases, information redundancy also increases. The redundancy not only influences the computational cost, but also reduces the precision of the regression model.

Through the PCA algorithm, the reconstructed temperature variables $T_1' \sim T_{20}'$ have been obtained. Here LGBM is used for selecting the most effective temperature features. The information gain of each temperature characteristic is calculated by the LGBM algorithm. Select temperature variables in order of information gain from high to low until the sum of the information gains reaches 95%. The selected temperature variables contain the optimal information abundance and are used as the input features for building the regression model.

Extraction Process of the Temperature Features via PCA-LGBM Data from the health monitoring system between January 2020 and December 2020 are used. $T_1' \sim T_{20}'$ are the reconstructed temperature variables via PCA. The information gains of $T_1' \sim T_{20}'$ are calculated via LGBM. Fig. 4 shows the ratio between the information gain and the total information gain of $T_1' \sim T_{20}'$. To retain the original information as much as possible while reducing the computational cost, T_4' , T_9' , T_{11}' , and T_{14}' , which include more than 95% of the total information gain, are selected as the input features of the regression model and are renamed T_a , T_b , T_c , and T_d . From January to December, each temperature feature retains 47,520 data points. Fig. 5 shows the time-history curves of the four temperature features.

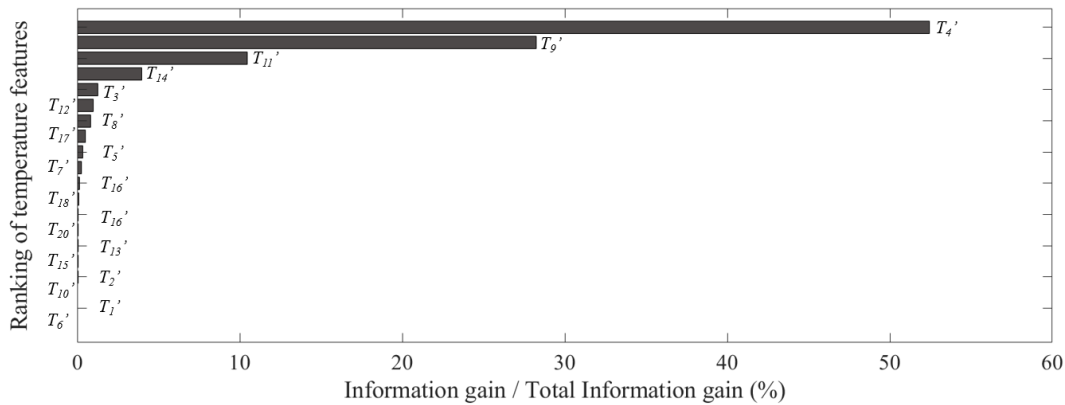
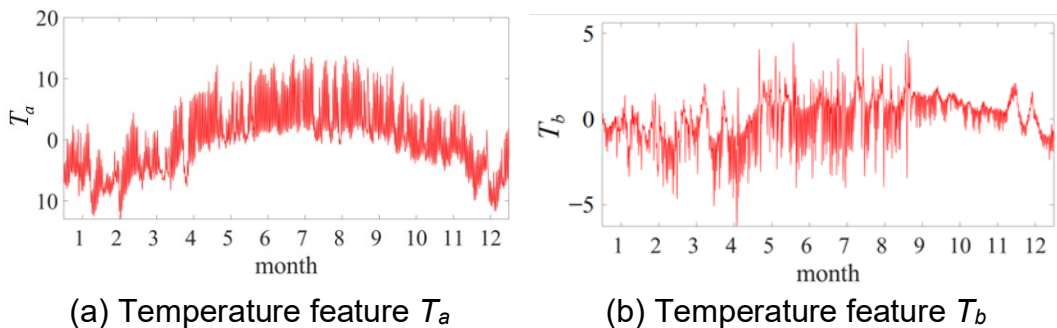


Fig. 4 The ratio of the information gain and the total information gain of each variable via LGBM



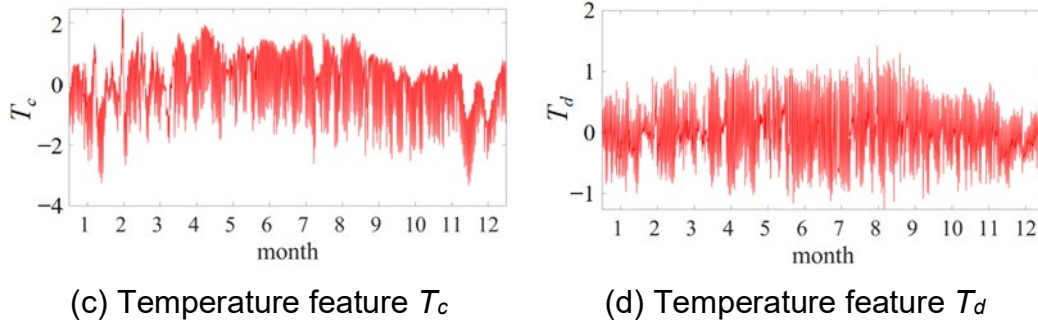


Fig. 5 Time–history curves of the four main temperature features

2.4 Constructing the Data Mode for Neural Networks

We could use the deflection data D and the temperature features T_a , T_b , T_c , and T_d to establish a regression model. To make the model have high precision and generalization performance, dividing the 47,520 data points into the training set contains 38,016 (80%) data points and the test set contains 9504 (20%) data points.

According to the above, we can form the input data as the dataset $X = \{x_1, x_2, \dots, x_t\}$. The x_t is described by Eq. (8).

$$x_t = [T_{a(t)} \ T_{b(t)} \ T_{c(t)} \ T_{d(t)}] \quad (8)$$

According to the existing research result, when building the mapping model of temperature and temperature-induced response via neural networks, using the temperature data of the first five hours as input data and the response data of the current moment as output data, the model will obtain the highest generalization performance. (Yue et al. 2022) Therefore, the data flow shown in Fig. 6 is constructed as the input–output mode for establishing the regression model. When the deflection D_t at time t is needed to be fitted, the thirty arrays of the current and previous moments from the dataset X are input. With the passage of time, this mapping mode also constantly changes as shown in Fig. 6.

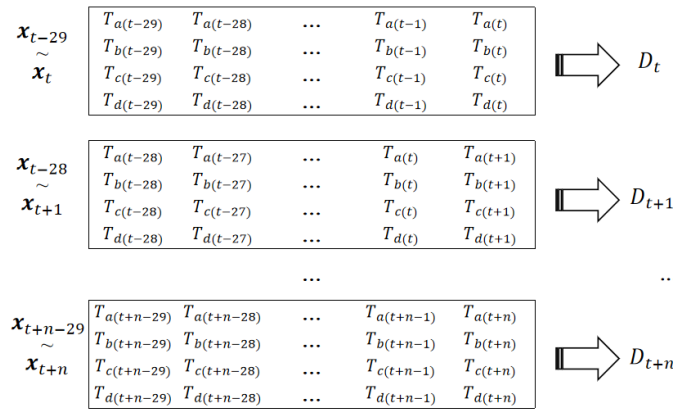


Fig. 6 Mapping mode between the temperature features and the temperature response

3. BACKPROPAGATION NEURAL NETWORK (BPNN)

BPNN is a widely used machine learning method with excellent performance in digital regression. A standard BPNN usually consists of an input layer, a hidden layer, and an output layer. (Liu et al. 2016) As shown in Fig. 7, the BPNN has two hidden layers and each layer has m units as an example. In the forward propagation of the BPNN, the data features from 1 to t are firstly input into the input layer. After the calculation using several weight coefficients W_{tm} , the calculated information is transferred to the hidden layer. After the operation via the activation function σ , the information is processed via the weighting coefficients W_{ik}^1 in the first hidden layer. Then, the information is further input into the next hidden layer and calculated via the activation function and the weighting coefficients W_k^2 in the second hidden layer. Finally, the fully connected layer integrates the values from the second hidden layer and thereby outputs the predicted regression value y' . In the training process, the parameters of the BPNN are adjusted and optimized through the backpropagation algorithm to improve the precision of the digital regression model.

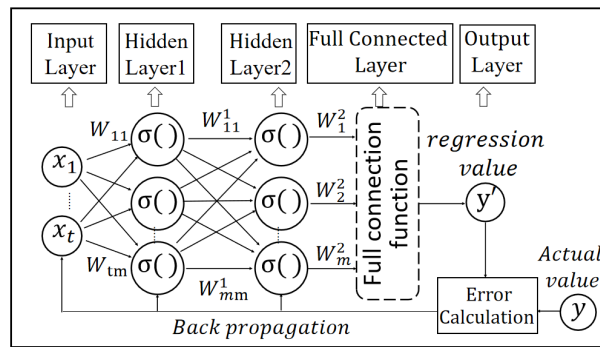


Fig. 7 The BPNN architecture with double hidden layers

BPNN is the most classic neural network, and it is used for preliminary exploratory modeling and serves as a benchmark for evaluating other subsequent methods. In this paper, the regression models between the temperature features and the temperature-induced deflection were established via BPNN using the s-type transfer function (logsig function) as the activation function.

To find the optimal BPNN model parameters, the different BPNN models are trained and tested. The BPNN models have 1 to 5 hidden layers, with each layer having 64, 128, or 256 hidden units. In the training phase, 0.1 is set as the initial learning rate. Fig. 8 shows that the root mean square error (RMSE) of the training set changes during the 500 epochs and iterates training of the BPNN model with different architectures. The RMSE is described by Eq. (9):

$$RMSE = \sqrt{\frac{\sum_{n=1}^N (D_n' - \bar{D})^2}{N}} \quad (9)$$

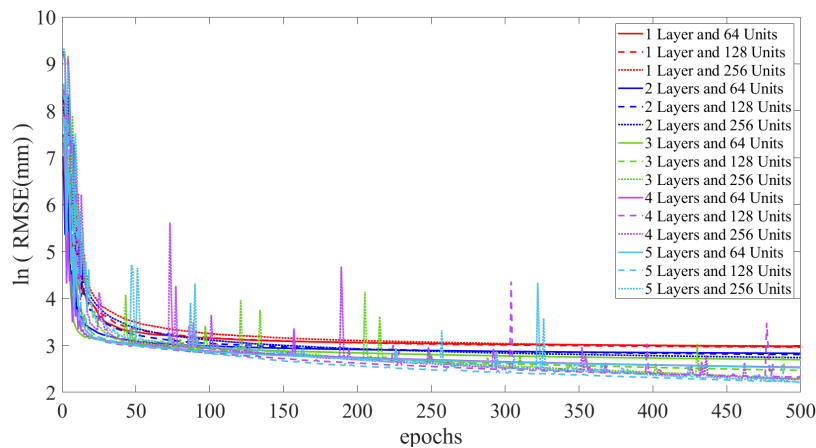


Fig. 8 RMSE changes during the training process of BPNN with different architectures

As shown in **Fig. 8**, when the number of the hidden layers is one, two, or three, the precision constantly increases. When the hidden layers are increased to four or five layers, the precision almost no longer improves, and the performances of BPNN with 128 hidden units and 256 hidden units in each hidden layer are not much different. Therefore, under 500 epochs of iteration, the network architecture with five hidden layers and 256 hidden units in each layer is selected to build the regression model.

Fig. 9 shows the output results of the regression model established via the BPNN with 5 hidden layers and 256 hidden units in each hidden layer. The *MSE* of the BPNN model was reduced to 154.09 mm², but this level of precision is obviously still insufficient. The mean square error (*MSE*) is described by **Eq. (10)**.

$$MSE = \frac{\sum_{n=1}^N (D_n' - \bar{D})^2}{N} \quad (10)$$

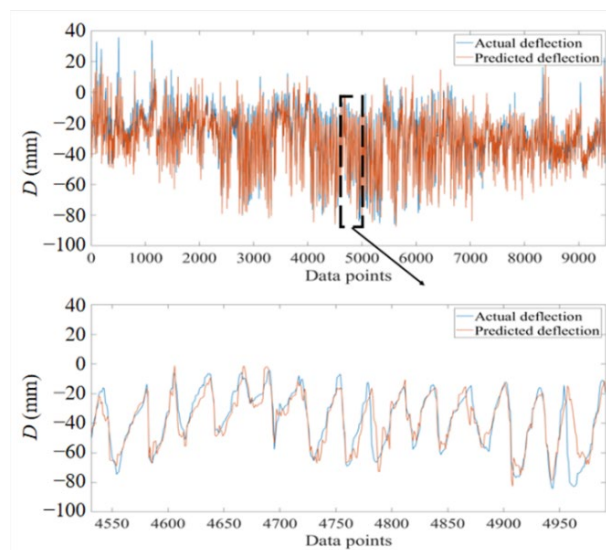


Fig. 9 Curves of the predicted deflection via the BPNN model

4. REGRESSION MODEL BASED ON DEEP LEARNING

The deep learning network has deeper hidden layers and a more complex computing unit in the hidden layer. It usually has a stronger nonlinear fitting performance than traditional networks, resulting in higher precision in regression modeling. (Williams & Zipser 1989; Yue et al. 2021) Therefore, the deep learning networks are used to further improve the regression performance in this paper.

4.1 Long-Short Term Memory Network (LSTM)

The recurrent neural network (RNN) usually performs well in modeling the time-varying correlation. (Hang et al. 2019) Compared with BPNN, RNN considers the information at the current moment and the previous information. Therefore, in this study a time series is used as the input data in the RNN regression model, and then the fitted data at the current moment are output.

The LSTM network is one type of the improved RNNs. LSTM can more accurately describe temporal nonlinear factors. It can fully mine the longer time series data by its specific gate functions in the unit of the hidden layer. (Zhang et al. 2018) Therefore, the LSTM network satisfies the requirements of explaining the nonlinear time dependence between the different variables. Fig. 10 shows the network architecture of LSTM with one LSTM hidden layer.

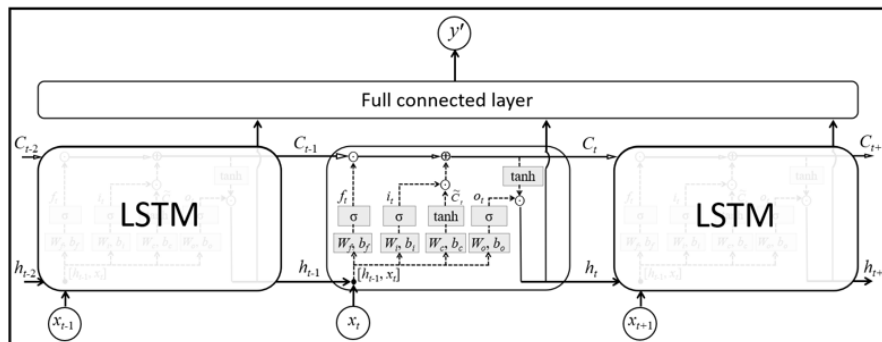


Fig. 10 LSTM network architecture with one LSTM hidden layer

The data flow in the hidden units and the hidden layers of the LSTM network is more complex than that of the RNN network. For a unit of the LSTM hidden layer, except putting the time series data $\{x_1, x_2, \dots, x_t\}$ into the model, the input data also include the information h_{t-1} transmitted from the LSTM unit at the previous moment, and includes the long-term memory C_{t-1} flowing in the cell. Several weights, biases, and activation functions constitute the forget gate, input gate, and output gate. Among the parameters in the LSTM network, f_t is generated from the forget gate, i_t and c_t are generated from the input gate, and O_t is generated from the output gate. The gate parameters are calculated by Eq. (11) to Eq. (14):

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (11)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (12)$$

$$\bar{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c) \quad (13)$$

$$O_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (14)$$

where W_c, W_f, W_i, W_o are weight coefficients; b_c, b_f, b_i, b_o are biases.

Due to the existence of the forget gate, the long-term information transmitted by the previous cell is only partially retained. The product of the parameters f_t and C_{t-1} , and the product of the parameters \bar{i}_t and \bar{C}_t , determine the update of information. Therefore, the current C_t can be updated to achieve better computational results. The parameter C_t is described by Eq. (15):

$$C_t = f_t \cdot C_{t-1} + \bar{i}_t \cdot \bar{C}_t \quad (15)$$

h_t is not only the output value of the network at the current moment, but also the information passed to the next cell. h_t is described by Eq. (16):

$$h_t = O_t \cdot \tanh(C_t) \quad (16)$$

4.2 Nested Long Short-Term Memory Network (NLSTM)

NLSTM is an optimized LSTM architecture jointly proposed by Carnegie Mellon University and the University of Montreal. NLSTM can express a higher nonlinear relationship than traditional LSTM by the more refined calculation in the hidden units.

As shown in Fig. 11, the NLSTM unit has a deeper architecture than the LSTM unit. The parameters in NLSTM can be described by Eq. (12) to Eq. (14) and Eq. (17) to Eq. (24):

$$\tilde{x}_t = i_t \times \sigma(W_c[h_{t-1}, x_t] + b_c) \quad (17)$$

$$\tilde{h}_{t-1} = f_t \cdot C_{t-1} \quad (18)$$

$$\bar{f}_t = \sigma(\bar{W}_f[\tilde{h}_{t-1}, \tilde{x}_t] + \tilde{b}_f) \quad (19)$$

$$\tilde{i}_t = \sigma(\bar{W}_i[\tilde{h}_{t-1}, \tilde{x}_t] + \tilde{b}_i) \quad (20)$$

$$\bar{O}_t = \sigma(\bar{W}_o[\tilde{h}_{t-1}, \tilde{x}_t] + \tilde{b}_o) \quad (21)$$

$$\bar{C}_t = \bar{f}_t \cdot \bar{C}_{t-1} + \tilde{i}_t \cdot \sigma(\bar{W}_c[\tilde{h}_{t-1}, \tilde{x}_t] + \tilde{b}_c) \quad (22)$$

$$C_t = O_t \cdot \sigma(C_t) \quad (23)$$

$$h_t = O_t \cdot C_t \quad (24)$$

where W_C , W_f , W_i , W_o are weight coefficients; b_C , b_f , b_i , b_o are biases.

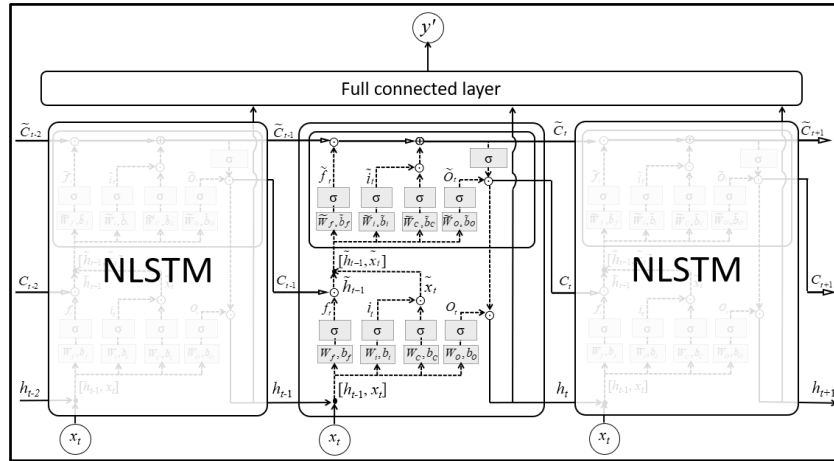


Fig. 11 Architecture of NLSTM network with one hidden layer

4.3 Optimized Architecture for NLSTM Network

The NLSTM network is established based on MATLAB. It also establishes the regression relationship between the temperature features changing over time and D_t .

Because the dataset was processed via the PCA-LGBM algorithm, the data do not need to be normalized. In the NLSTM network, the training datasets are still X_1 and D_1 . The NLSTM network and its parameters were dissected as shown in Fig. 11. During the training phase, the parameters are continuously adjusted via backpropagation.

After the NLSTM network undergoes the preset iterative epochs of optimization training, a regression model is obtained. To test the performance of this regression model, the test set X_2 is used as the input data, and the predicted values calculated by the model are compared with the actual D_2 .

Due to the limited capacity of the NLSTM with only one hidden layer for handling engineering problems, the neural networks must be organized into an optimized architecture. (Yu et al. 2019; Nakisa et al. 2018) It is necessary to find the best network architecture for NLSTM to ensure accuracy.

Considering the data scale and the computing power of our device, this paper sets the batch size as 32, and uses the Adam optimizer in backward propagation. The training process has 100 iterative epochs. To speed up the convergence speed, the initial learning rate is set to 0.001, and it will reduce to 0.0002 after 50 epochs to ensure the convergence of the NLSTM network.

The datasets X_1 and D_1 are utilized to, respectively, train NLSTM networks with single, double, and triple hidden layers. The units in each hidden layer change from 64 hidden units to 128 hidden units to 256 hidden units. Fig. 12 shows the root mean square

error (*RMSE*) of the training set changes during 100 epochs of nine NLSTM networks with different architectures.

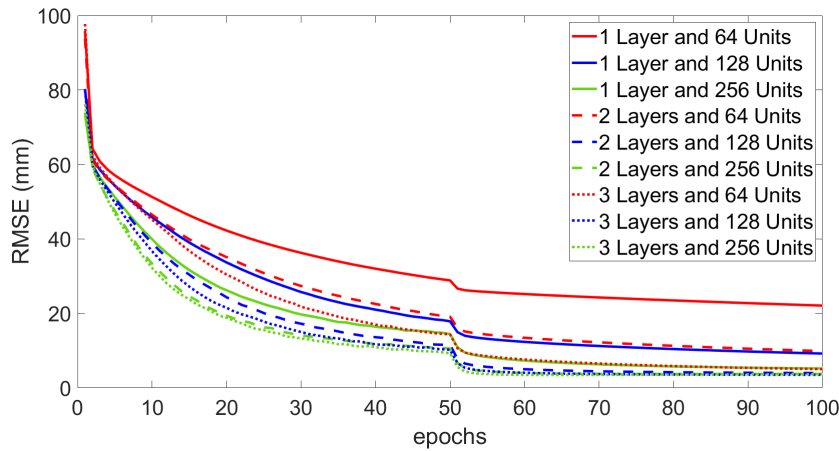


Fig. 12 *RMSE* changes during the training process of the NLSTM networks

The test sets X_2 and D_2 were applied to test the precision of the nine models after the training. **Fig. 13** shows the mean square error of the predicted results of the nine NLSTM networks with different architectures.

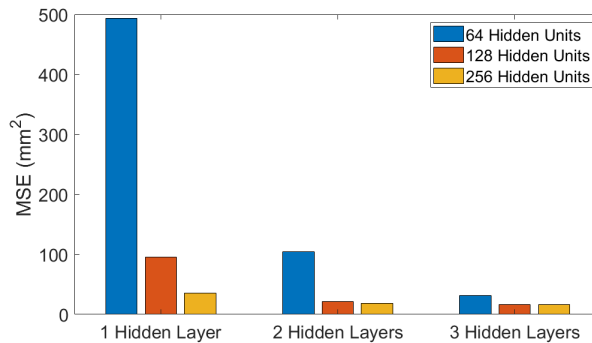


Fig. 13 *MSE* of the predicted results of the test set via different NLSTM networks

Benefiting from the excellent temperature feature extracted via the PCA-LGBM algorithm and reasonable hyperparameter selection, even the NLSTM network with 1 hidden layer and 64 hidden units does not show a significant difference between the training results and testing results. This proves that there is no overfitting or other singular phenomenon when the trained model is used for the practical test.

The *RMSE* of the NLSTM network with n hidden layers and m hidden units after 100 epochs of training is represented as $R_{n,m}$. The root mean square errors of different NLSTM networks are sorted from largest to smallest as shown below:

$$R_{1,64} > R_{2,64} > R_{1,128} > R_{1,256} > R_{3,64} > R_{2,128} > R_{2,256} > R_{3,128} > R_{3,256}$$

In considering the training effect alone, it can be seen that all nine networks are

converged. The root mean square error of the NLSTM network with a single hidden layer and 64 hidden units is larger than other networks. Of course, an evaluation of the ability of a neural network must also consider the test effect.

For the test results, a similar conclusion can also be proved from the mean square errors of the nine NLSTM networks in Fig. 13. The root mean square error reduces the increases in the number of hidden layers and hidden units in the NLSTM network. The *MSE* of some NLSTM networks are very similar. This illustrates that there is no significant difference in fitting ability between the NLSTM networks, and the networks should be further selected by considering the computational efficiency.

We can further evaluate the performances and efficiencies of the NLSTM networks with different architectures through the test precision and training time. After 100 epochs of iterative training, for the nine different NLSTM networks, the training time and the *MSE* of the test set make up the two arrays as follows:

$$\{t_{1_{64}}, t_{1_{128}}, t_{1_{256}}, t_{2_{64}}, t_{2_{128}}, t_{2_{256}}, t_{3_{64}}, t_{3_{128}}, t_{3_{256}}\}$$

$$\{M_{1_{64}}, M_{1_{128}}, M_{1_{256}}, M_{2_{64}}, M_{2_{128}}, M_{2_{256}}, M_{3_{64}}, M_{3_{128}}, M_{3_{256}}\}$$

where t_{n_m} represents the time cost in 100 epochs of training for the NLSTM network with n hidden layers and each hidden layer with m hidden units. M_{n_m} represents the *MSE* of the test set of the NLSTM network with n hidden layers and each layer with m hidden units. η_{n_m} is the product of the corresponding values in the two arrays. The parameter η_{n_m} is described by Eq. (25):

$$\eta_{n_m} = \frac{1}{t_{n_m}} \times \frac{1}{M_{n_m}} \quad (25)$$

Fig. 14 shows the comparison of the η_{n_m} of the different NLSTM networks. The NLSTM networks with 64 hidden units per hidden layer are the most inefficient. The NLSTM network with 2 hidden layers and each layer with 256 hidden units is the optimal choice.

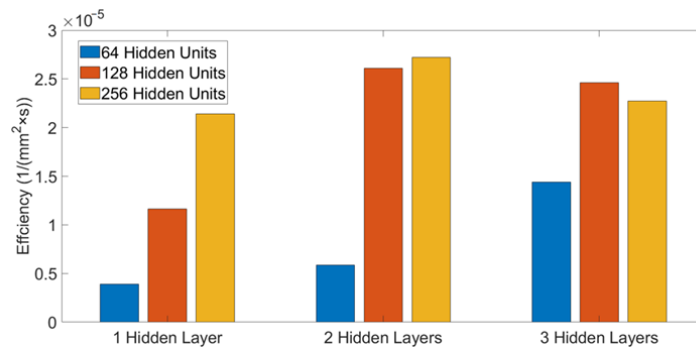


Fig. 14 Efficiencies of the different NLSTM networks

Fig. 15 shows the output results of the regression model established by the NLSTM network which has 2 hidden layers and 256 hidden units in each hidden layer.

The mean square error of the test set is 18.57 mm^2 and it is more accurate than the BPNN regression model.

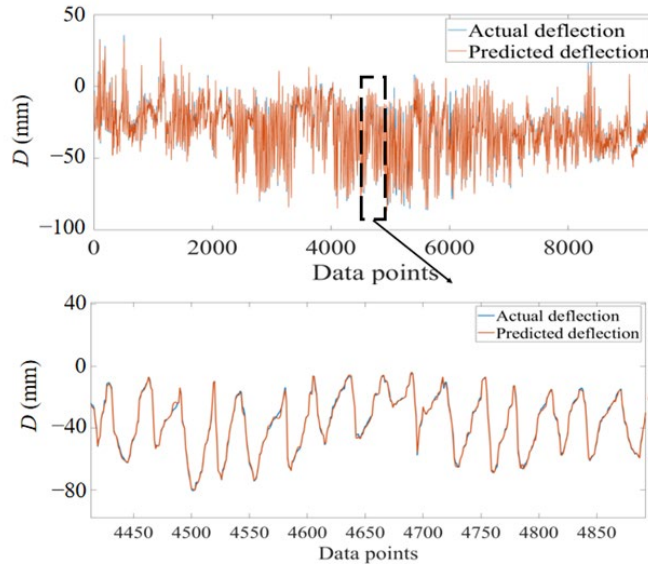


Fig. 15 Curves of the predicted deflection obtained via the NLSTM model

To demonstrate that the NLSTM unit has better performance than the classical LSTM unit, the LSTM regression model is also established with 2 hidden layers and 256 hidden units for the test set. For this LSTM model, the *MSE* of the test set is 66.95 mm^2 . **Fig. 16** shows the output result of the LSTM regression model. Obviously, the precision of NLSTM is superior to that of LSTM. The results of different models are further compared and discussed in the next section.

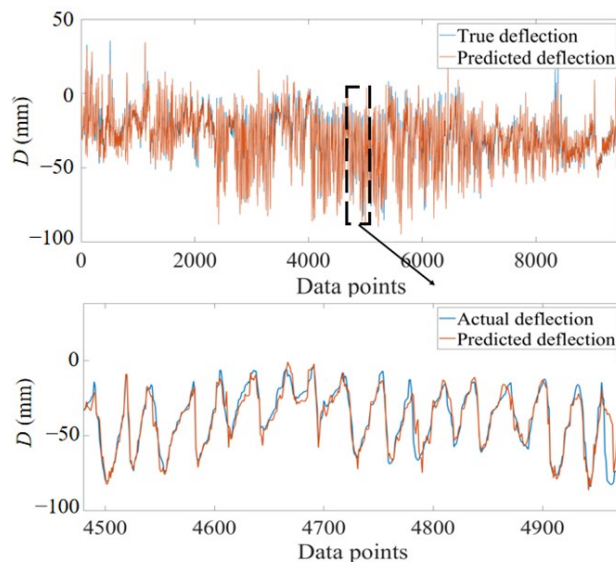


Fig. 16 Curves of the predicted deflection obtained via the LSTM model

5. EVALUATION OF SEVERAL MODELS

Three indicators were selected to assess the accuracy of different models more comprehensively. These three indicators are, respectively, the *MSE*, which is described by Eq. (10), the Maximum Absolute Error (*MAXAE*), described by Eq. (26), and the Mean Absolute Error (*MAE*), described by Eq. (27). The *MAE* evaluates the overall error; the *MSE* evaluates the output stability; and the *MAXAE* evaluates the extreme value of error.

$$MAXAE = \max |D_n - D_n'| \quad (26)$$

$$MAE = \frac{1}{n} \sum |D_n - D_n'| \quad (27)$$

The above error indicators were, respectively, tested using the regression values from the BPNN, LSTM, and NLSTM regression models which were trained using the datasets from the combined PCA-LGBM. To demonstrate the applicability and effectiveness of the four extracted temperature features of the combined PCA-LGBM algorithm, the same training and testing procedures were performed by randomly selecting four temperature features, and the above error indicators of the selected methods of the randomly selected four temperature features are calculated again. The relevant calculation results are shown in Table 1.

Table 1 *MAE*, *MSE*, and *MAXAE* of different methods

Methods	<i>MAE</i> (mm)	<i>MSE</i> (mm ²)	<i>MAXAE</i> (mm)
BPNN (features of PCA-LGBM)	12.63	154.09	50.44
LSTM (features of PCA-LGBM)	8.06	66.95	33.16
NLSTM (features of PCA-LGBM)	4.76	18.57	27.37
BPNN (randomly selected features)	19.55	326.78	62.89
LSTM (randomly selected features)	18.86	267.86	56.87
NLSTM (randomly selected features)	17.34	250.16	47.98

The discussion of the several calculation results is as follows:

When the regression model is established using the NLSTM network with 2 hidden layers and 256 hidden units in each hidden layer, for the fitting with features via PCA-LGBM, the model accuracy and stability were significantly improved.

When using the randomly selected features, the errors of the calculation results obtained using the selected methods are all too large for engineering application. Obviously, if the extracted datasets obtained using the PCA-LGBM intelligent algorithm

presented in this paper are not used, the reliable model cannot be established even by using the advanced fitting tools.

Therefore, the regression model established using the NLSTM network has superior stability and accuracy with appropriate hyperparameters and reasonable datasets extracted via PCA-LGBM. The model demonstrates a very high precision and can, therefore, provide a very valuable control group for bridge maintenance. Abnormal bridge states can be detected as early as possible; therefore, the model will play a huge role in the assessment of bridge states.

6. CONCLUSIONS

The relevant conclusions of this article are as follows:

- (1) After verification, for the structural health monitoring data, the main temperature features extracted from the complex temperature field via the PCA-LGBM algorithm have a reliable generalization. It is impossible to build a precise model using the randomly selected features. Thus, the algorithm plays an important role in intelligent regression analysis.
- (2) The architecture of the NLSTM network should be optimized. On the one hand, if there are too few hidden layers and units, the requirements of prediction accuracy and generalization performance will be difficult to satisfy. On the other hand, if the layers and units are too numerous, a substantial computational cost will be incurred with little improvement in model accuracy. In this regard, the NLSTM network with 2 hidden layers, each with 256 hidden units, is the best choice.
- (3) The deep learning-based NLSTM network has higher stability and accuracy in regression analysis compared with machine learning-based BPNN. Its *MSE* is only one-eighth of that of the BPNN regression model, illustrating the strong stability of the pro-posed NLSTM model. Similarly, the calculation results of NLSTM are better than those of classical LSTM, indicating that using NLSTM to build the regression model of the temperature-induced deflection for cable-stayed bridges is preferred. The significance and research prospects of this article are as follows:

By using NLSTM and the features extracted via PCA-LGBM, the model can reach a very high precision, thus providing a very valuable control group for bridge maintenance. Therefore, the abnormal state of the bridge can be detected as early as possible. Of course, this paper only provides one fitting tool, and using the regression value of this tool to detect abnormal states still requires the further work consisting of numerical simulations, experiments, and various mathematical tools. This research will continue in the future.

REFERENCES

- Cai, G.W. and Mahadevan, S. (2018), "Big Data Analytics in Uncertainty Quantification: Application to Structural Diagnosis and Prognosis", *Asce-Asme Journal of Risk and Uncertainty in Engineering Systems Part a-Civil Engineering*, 4(1).

- Ding, Y., Wang, G., Zhou, G. and Li, A. (2013), "Life-cycle simulation method of temperature field of steel box girder for Runyang cable-stayed bridge based on field monitoring data", *China Civil Engineering Journal*, **46**(5), 129-136.
- Erichson, N.B., Zheng, P., Manohar, K., Brunton, S.L., Kutz, J.N. and Aravkin, A.Y. (2020), "SPARSE PRINCIPAL COMPONENT ANALYSIS VIA VARIABLE PROJECTION", *Siam Journal on Applied Mathematics*, **80**(2), 977-1002.
- Hang, R.L., Liu, Q.S., Hong, D.F. and Ghamisi, P. (2019), "Cascaded Recurrent Neural Networks for Hyperspectral Image Classification", *Ieee Transactions on Geoscience and Remote Sensing*, **57**(8), 5384-5394.
- Huang, H.B., Yi, T.H. and Li, H.N. (2016), "Canonical correlation analysis based fault diagnosis method for structural monitoring sensor networks", *Smart Structures and Systems*, **17**(6), 1031-1053.
- Ju, Y., Sun, G.Y., Chen, Q.H. Zhang, M., Zhu, H.X. and Rehman, M.U. (2019), "A Model Combining Convolutional Neural Network and LightGBM Algorithm for Ultra-Short-Term Wind Power Forecasting", *Ieee Access*, **7**, 28309-28318.
- Liu, Y., Jing, W.Z. and Xu, L.X. (2016), *Cascading Model based Back Propagation Neural Network in Enabling Precise Classification*, 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Changsha, PEOPLES R CHINA.
- Ma, X.L., Zhong, H.Y., Li, Y., Ma, J.Y., Cui, Z.Y. and Wang, Y.H. (2021), "Forecasting Transportation Network Speed Using Deep Capsule Networks With Nested LSTM Models", *Ieee Transactions on Intelligent Transportation Systems*, **22**(8), 4813-4824.
- Miyamoto, A. and Motoshita, M. (2017), "A Study on the intelligent bridge with an advanced monitoring system and smart control techniques", *Smart Struct. Syst*, **19**, 587-599.
- Nakisa, B., Rastgoo, M.N., Rakotonirainy, A., Maire, F. and Chandran, V. (2018), "Long Short Term Memory Hyperparameter Optimization for a Neural Network Based Emotion Recognition Framework", *Ieee Access*, **6**, 49325-49338.
- Punmiya, R. and Choe, S. (2019), "Energy Theft Detection Using Gradient Boosting Theft Detector With Feature Engineering-Based Preprocessing", *Ieee Transactions on Smart Grid*, **10**(2), 2326-2329.
- Singh, A., Kushwaha, S., Alarfaj, M. and Singh, M. (2022), "Comprehensive Overview of Backpropagation Algorithm for Digital Image Denoising", *Electronics*, **11**(10).
- Smagulova, K. and James, A.P. (2019), "A survey on LSTM memristive neural network architectures and applications", *European Physical Journal-Special Topics*, **228**(10), 2313-2324.
- Sun, L., Shang, Z., Xia, Y., Bhowmick, S. and Nagarajaiah, S. (2020), "Review of Bridge Structural Health Monitoring Aided by Big Data and Artificial Intelligence: From Condition Assessment to Damage Detection", *Journal of Structural Engineering*, **146**(5), 04020073.
- Tu, C., Liu, Z., Zhang, G., Zhou, L., Chen, Y., Cheng, N., Gu, J., Dong, S., Deng, Z., Wang, Y. and Tang, L. (2017), "Processing Technique and Application of Big Data Oriented to Long-term Bridge Health Monitoring", *Journal of Experimental Mechanics*, **32**(5), 652-663.

- Wang, L., Zeng, Y. and Chen, T. (2015), "Back propagation neural network with adaptive differential evolution algorithm for time series forecasting", *Expert Systems with Applications*, **42**(2), 855-863.
- Wang, M.Y., Ding, Y.L., Wan, C.F. and Zhao, H.W. (2020), "Big data platform for health monitoring systems of multiple bridges", *Structural Monitoring and Maintenance*, **7**(4), 345-365.
- Williams, R.J. and Zipser, D. (1989), "A Learning Algorithm for Continually Running Fully Recurrent Neural Networks", *Neural Computation*, **1**(2), 270-280.
- Xia, G.P. (2014), "Parametric Study of Cable Deflection and Gravity Stiffness of Cable-Stayed Suspension Bridge", *Applied Mechanics and Materials*, **488-489**, 445-448.
- Xu, X., Xu, C.B., Zhang, Y. and Wang, H.L. (2021), "Preliminary Study on the Loss Laws of Bearing Capacity of Tunnel Structure", *Symmetry-Basel*, **13**(10).
- Yi, Z., Sun, L.M., Fu, Z.H. and Jiang, Z. (2020), "Study on temperature sensitivity coefficients of mid-span vertical displacement of cable-stayed bridges", *Eng. Mech.*, **37**(6), 148-154. (In Chinese)
- Yu, Y., Si, X.S., Hu, C.H. and Zhang, J.X. (2019), "A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures", *Neural Computation*, **31**(7), 1235-1270.
- Yue, Z.X., Ding, Y.L., Zhao, H.W. and Wang, Z.W. (2021), "Case Study of Deep Learning Model of Temperature-Induced Deflection of a Cable-Stayed Bridge Driven by Data Knowledge", *Symmetry-Basel*, **13**(12).
- Yue, Z.X., Ding, Y.L., Zhao, H.W. and Wang, Z.W. (2022), "Mechanics-Guided optimization of an LSTM network for Real-Time modeling of Temperature-Induced deflection of a Cable-Stayed bridge", *Engineering Structures*, **252**, 113619.
- Zhang, Y.Z., Xiong, R., He, H.W. and Pecht, M.G. (2018), "Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries", *Ieee Transactions on Vehicular Technology*, **67**(7), 5695-5705.
- Zhao, H.W., Ding, Y.L., Nagarajaiah, S. and Li, A.Q. (2019), "Behavior Analysis and Early Warning of Girder Deflections of a Steel-Truss Arch Railway Bridge under the Effects of Temperature and Trains: Case Study", *Journal of Bridge Engineering*, **24**(1), 05018013.
- Zhou, Y. and Sun, L. (2019), "Insights into temperature effects on structural deformation of a cable-stayed bridge based on structural health monitoring", *Structural Health Monitoring*, **18**(3), 778-791.
- Zhou, Y., Sun, L.M., Fu, Z.H., Jiang, Z. and Ren, P.J. (2020), "General formulas for estimating temperature-induced mid-span vertical displacement of cable-stayed bridges", *Engineering Structures*, **221**.