Smart Structures and Systems, *Vol. 29, No. 1 (2022) 117-127*
https://doi.org/10.12989/sss.2022.29.1.117

Data abnormal detection using bidirectional long-short neural network combined with artificial experience

Kang Yang [a†], Huachen Jiang [b†], Youliang Ding[*], Manya Wang [c] and Chunfeng Wan [d]

School of Civil Engineering, Key Laboratory of C&PC Structures of the Ministry of Education, Southeast University, Nanjing 210096, China

(*Received April 14, 2021, Revised June 12, 2021, Accepted July 12, 2021*)

Abstract. Data anomalies seriously threaten the reliability of the bridge structural health monitoring system and may trigger system misjudgment. To overcome the above problem, an efficient and accurate data anomaly detection method is desiderated. Traditional anomaly detection methods extract various abnormal features as the key indicators to identify data anomalies. Then set thresholds artificially for various features to identify specific anomalies, which is the artificial experience method. However, limited by the poor generalization ability among sensors, this method often leads to high labor costs. Another approach to anomaly detection is a data-driven approach based on machine learning methods. Among these, the bidirectional long-short memory neural network (BiLSTM), as an effective classification method, excels at finding complex relationships in multivariate time series data. However, training unprocessed original signals often leads to low computation efficiency and poor convergence, for lacking appropriate feature selection. Therefore, this article combines the advantages of the two methods by proposing a deep learning method with manual experience statistical features fed into it. Experimental comparative studies illustrate that the BiLSTM model with appropriate feature input has an accuracy rate of over 87-94%. Meanwhile, this paper provides basic principles of data cleaning and discusses the typical features of various anomalies. Furthermore, the optimization strategies of the feature space selection based on artificial experience are also highlighted.

Keywords: BiLSTM; data anomaly detection; feature extraction; long-span bridge; structural health monitoring

1. Introduction

Thanks to the rapid development of big data technology, data has increasingly become the core of improving efficiency in all walks of life. As to the bridge health monitoring, monitoring data is no exception: considering that data underlies structural evaluation and abnormal warning (Bao *et al*. 2019a, Bao and Li 2020). Local damages are identified by detecting abnormal changes in the measured data (Bao *et al*. 2019b, Liu *et al*. 2020). However, the causes of sensor failures and structural damage are coupled and will trigger false alarms, as the former sometimes obscures the latter. As a result, developing algorithms to identify abnormal data, determine whether it is a sensor failure, and eliminate its impact on subsequent evaluation is crucial to ensure and prolong the normal operation of the SHM system.

So far, in the data anomaly detection stage, algorithms can be divided into two major categories: statistical model-based and data-driven approaches (Yu *et al*. 2013, Chen *et al*. 2018a). In the model-based approaches, knowledge of statistical and physical models on the measured data is taken to form a mathematical model. In these methods, a model is initially constructed to describe normality through non-faulty health monitoring data. According to different usage of models, these traditional model-based methods can be divided into methods that are based on univariate control charts and multivariate statistical analysis (Yi *et al*. 2017, Chen *et al*. 2018b, 2019). Particularly, the method based on the univariate control chart focuses on detecting abnormalities in a single sensor. However, the method based on multivariate statistical analysis has taken into consideration correlation issues, so it outperforms in the fault detection field compared with the method based on the univariate control chart. Once anomalies are found through multivariate statistical analysis, several fault isolation methods can be used to further isolate faulty sensors, such as contribution analysis, missing variable method, and probability quantification (Huang *et al*. 2016, Kullaa 2013, Ni *et al*. 2019, Tang *et al*. 2019, Xia and Ni 2018, Yang and Nagarajaiah 2016). The model-based approach can achieve high recognition accuracy by refined modeling and targeted study of anomaly types, but it has poor applicability for sensors in different locations and environments, which brings the drawback of labor-intensive. Data-driven methods now mainly adopted artificial intelligence techniques (e.g., machine learning and deep learning) for the classification task. Liu *et al*. (2004) adopted the unbalanced binary tree method to categorize the type of

*Corresponding author, Ph.D., Professor,
E-mail: civilchina@hotmail.com
† Kang Yang and Huachen Jiang contribute equally to this article, they are the co-first authors.
[a] Ph.D. Student
[b] Ph.D. Student
[c] Ph.D. Student
[d] Ph.D.

ISSN: 1738-1584 (Print), 1738-1991 (Online)

sensor fault. Li *et al*. (2019) obtained statistical correlation coefficients and trained a relevance vector machine to classify sensor fault types. In addition to the above-mentioned machine learning algorithms based on time-domain features, some scholars transformed data into two-dimensional images and classified the sensor faults utilizing convolutional neural networks (CNN) (Spencer *et al*. 2019). Bao *et al*. (2019) transformed time-series signals into image vectors and trained via techniques termed stacked autoencoders, achieving high accuracy for multi-pattern anomalies detection. Tang *et al*. (2019) proposed a novel data anomaly detection method based on a dual-information CNN that imitates human vision making the detection process readily scalable, faster, and more accurate. However, all of the above methods have disadvantages in that the network relies on suitable feature inputs. Proper feature input is a prerequisite for the model to work accurately.

Although many researchers have proposed a large variety of methods for data anomaly classification, there is still a lack of detailed discussion on the basic principles of the development in automatic anomaly recognition algorithms for engineering applications, the extraction of abnormal features, and the comparison of different methods. Besides, the detection method based on traditional statistical distribution features has clear discriminative indicators and principles, but sensors from different positions and sources exhibit poor applicability, which requires manual intervention. The deep learning method has a strong generalization ability for specific anomaly recognition and rarely needs manual intervention (Yuen and Ortiz 2017, Liu *et al*. 2020). While when it comes to feature selection, there is insufficient theory to support what feature should be extracted. Therefore, it is worthwhile to establish a recognition algorithm that combines the traditional statistical experience with artificial intelligence formation, that is to consider a method that detects abnormity efficiently based on deep learning and statistical experience.

This paper mainly combines artificial experience and a deep learning algorithm based on BiLSTM to detect data abnormalities. The structure of this paper is illustrated in Fig. 1. First, according to the given tags, various types of abnormal data are analyzed statistically, and its time domain extracted. Then, a deterministic recognition algorithm based on the above steps is established. After adopting the general feature extraction method, a BiLSTM model was established under supervised training. The accuracy and applicability of the two methods are also verified and illustrated. Finally, from the engineering perspective, this paper concludes the general principles of monitoring data cleaning, the main characteristics of abnormal types, and the experience of the deep learning training process, all of the above are meaningful and instructive.

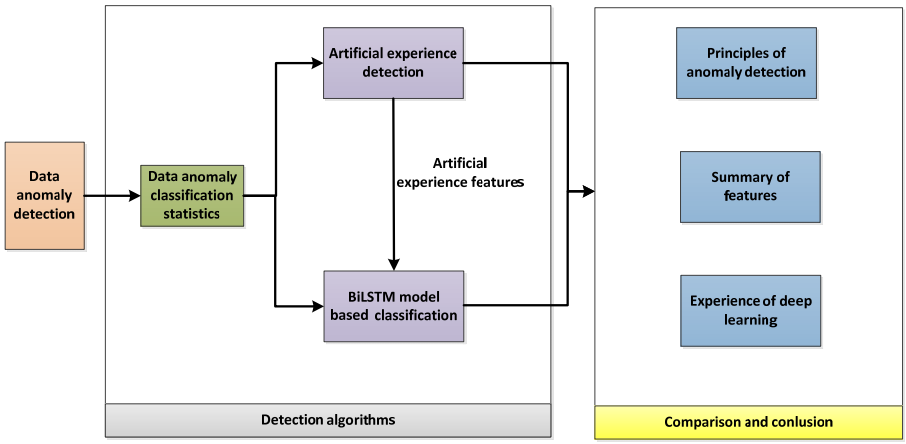


Fig. 1 Flow chart of this paper

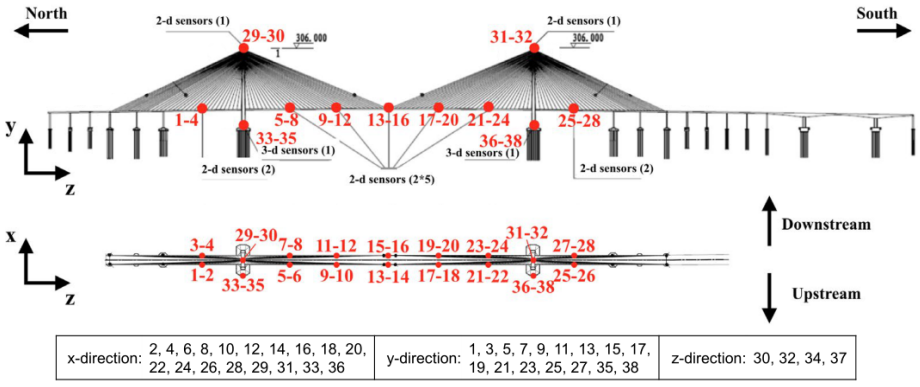


Fig. 2 The bridge and placement of accelerometers on the deck and towers

Table 1 Data pattern by sensors

	Classes of data pattern							
	1.Normal	2.Missing	3.Minor	4.Outlier	5.Square	6.Trend	7.Drift	Total
Sensor1	41	55	504	105	39	0	0	744
Sensor2	46	57	444	156	41	0	0	744
Sensor3	4	50	658	6	26	0	0	744
Sensor4	659	56	0	4	25	0	0	744
Sensor5	736	7	0	1	0	0	0	744
Sensor6	706	3	0	35	0	0	0	744
Sensor7	742	2	0	0	0	0	0	744
Sensor8	736	8	0	0	0	0	0	744
Sensor9	744	0	0	0	0	0	0	744
Sensor10	479	2	62	201	0	0	0	744
Sensor11	742	2	0	0	0	0	0	744
Sensor12	737	7	0	0	0	0	0	744
Sensor13	0	207	2	0	0	478	57	744
Sensor14	0	208	2	0	0	476	58	744
Sensor15	0	201	2	0	0	483	58	744
Sensor16	0	203	2	0	0	480	59	744
Sensor17	0	205	2	0	0	482	55	744
Sensor18	0	202	2	0	0	482	58	744
Sensor19	1	207	1	0	0	480	55	744
Sensor20	1	203	2	0	0	480	58	744
Sensor21	0	202	1	0	0	484	57	744
Sensor22	0	202	2	0	0	485	55	744
Sensor23	1	203	1	0	0	486	53	744
Sensor24	0	204	2	0	0	482	56	744
Sensor25	74	4	0	0	666	0	0	744
Sensor26	73	9	0	0	662	0	0	744
Sensor27	74	4	0	0	0	666	0	744
Sensor28	0	4	0	0	740	0	0	744
Sensor29	663	54	1	1	25	0	0	744
Sensor30	568	60	85	5	26	0	0	744
Sensor31	731	3	0	10	0	0	0	744
Sensor32	736	8	0	0	0	0	0	744
Sensor33	683	34	0	1	26	0	0	744
Sensor34	689	27	0	1	27	0	0	744
Sensor35	689	27	0	1	27	0	0	744
Sensor36	739	5	0	0	0	0	0	744
Sensor37	738	6	0	0	0	0	0	744
Sensor38	743	1	0	0	0	0	0	744

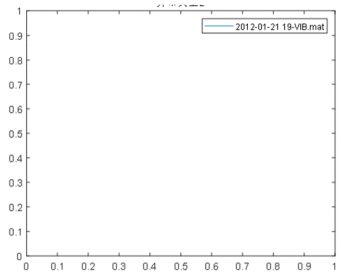
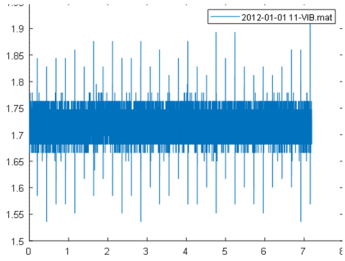
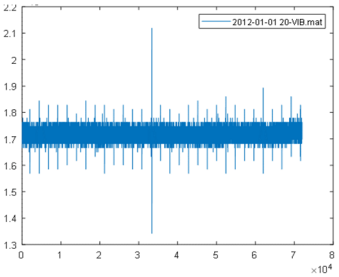
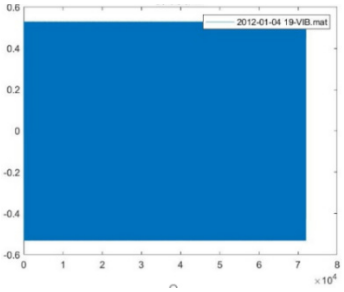
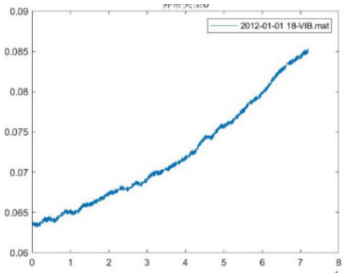
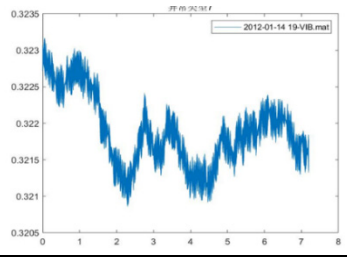
2. Artificial experience detection method

The overall idea of the anomaly detection method based on artificial experience contains the following steps. First, analyze the statistical characteristics of various types of anomalies by sensors. Then, extract the statistical features of various types of anomalies. At last, set thresholds and aggregate the thresholds of various types of anomalies for comprehensive anomaly detection.

2.1 Statistics of abnormal classes

The dataset consists of one month's acceleration data for a long-span cable-stayed bridge in China. There are 38 sensors, whose locations are illustrated in Fig. 2. The sampling frequency is 20 Hz. Additionally, a dataset is

Table 2 Description of the 6 abnormal data patterns

Data patterns	Typical representative	Description of the features: Description	Description of the features: The basis of artificial algorithms	Description of the features: Feature index for BiLSTM
2 Missing		1 A lot of vacant data; 2 Constant value	More than 90% of the data are NAN or a constant value	Mean, maximum values and variance of data per minute;
3 Minor		Most of the data is within a bandwidth	Most of the data is within a bandwidth and the maximum and minimum values of the difference index do not exceed the given threshold	Mean, maximum values and variance of data per minute;
4 Outlier		Few data points are significantly outlier	More than 85%of data is distributed within the bandwidth, and the maximum and minimum values of the difference index exceed the given threshold	Mean, maximum values and variance of data per minute;
5 Square		Strict oscillation repetition	All data oscillates within the bandwidth; values of the difference is the same	Entropy of data per minute
6 Trend		The data mean position has a significant trend	The fitting coefficient is greater than the given threshold	Mean and maximum values of data per minute
7 Drift		Data base position shows obvious fluctuations	There are multiple peaks at the reference position	Mean, maximum values and variance of data per minute;

labeled for the time series by the hour. For the one-month (31 days) of data for 38 sensors, the dimension of the labeled dataset is 744 × 38.

First, the labeled data set is classified and counted according to the sensor anomaly type to capture the distribution characteristics of various anomalies. As is shown in Table 1 where abnormal data are categorized. It can be found that different sensors tend to have certain types of data abnormalities. For example, sensors 1-4 only have the first four types of abnormalities, and sensors 13-24 basically only have type 2, 6, and 7. According to the types of abnormalities that occur by groups, the 38 sensors can be divided into 4 categories, as shown in the painted table parts. Besides, abnormal types 1-4 usually occur concurrently, meanwhile abnormal types 2, 6, and 7 also usually occur concurrently.

2.2 Feature extraction

By observing the characteristics of various abnormalities in the time domain, the decipherment of 6 abnormal data patterns is shown in Table 2. The table contains the descriptions of each abnormality, indicators of artificial algorithms, and features extracted for subsequent deep learning.

The feature thresholds of the manual empirical method are directly related to the accuracy of anomaly detection. The determination of thresholds consists of two parts: the identification of sensitive features for specific anomaly types, and the setting of feature thresholds. Firstly, nine common statistical features are selected, and each feature indicator is calculated for one hour of labeled data on a time scale of every minute, and then ranked by magnitude. The largest feature indicator corresponding to each type of anomaly is set to be the sensitive indicator for that type of anomaly. The feature threshold for each type of anomaly is then determined. The setting of the feature threshold is initially determined according to the 95% quantile of various features calculated on the data set. Then calculated the accuracy and recall rate, and fine-tuned the threshold of various features. So that the thresholds are set.

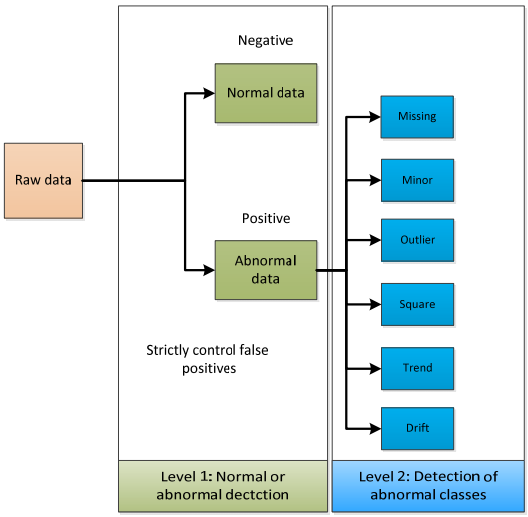


Fig. 3 Two-level recognition framework

2.3 A Two-levels recognition framework

Based on the engineering perspective, a two-level anomaly detection framework is proposed, as shown in Fig. 3. The framework first distinguishes whether the data is anomalous or not, and then classifies it by anomaly types. Considering the difficulty of repairing dynamic acceleration data, it is of great significance to determine if there is any abnormal in the first level.

2.4 Normal or abnormal data detection

One important trade-off in the classification problem is to balance false positives and false negatives. As for health monitoring data, to strictly control false negatives and appropriately relax false positives is highly suggested, because abnormal data can escape from data cleaning due to the wrong judgment by false negatives. When it is applied to subsequent data analysis, serious interference in the evaluation of the structural state will occur and cause a false early warning. Therefore, when setting and adjusting the algorithm threshold of the artificial method, cutting down the false-negative rate as much as possible is imperative. At the same time, the accuracy of identifying abnormal data and the false-positive and false-negative rate are given to show the achievement of the framework when the above principles are considered.

To quantify the prediction quality, common indicators are used to evaluate the classification problems. Among them, TPR, TNR, FPR, and FNR represent the true positive rate, true negative rate, false-positive rate, and false-negative rate respectively.

$$TPR = TP / (TP + FN) = TP / T \quad (1)$$

$$TNR = TN / (FP + TP) = TN / F \quad (2)$$

$$FPR = FP / (FP + TN) = FP/ F \quad (3)$$

$$FNR= FN / (TP + FN) = FN /T \quad (4)$$

$$Accuracy = (TP + TN)/(ALL) \quad (5)$$

where TP, FP, TN, and FN represent the number of true-positive samples, false-positive samples, true-negative samples, and false-negative samples.

By appropriately relaxing the threshold settings of various abnormal features, the false-negative rate has been controlled strictly below 0.5%. The first-level discrimination result of sensor 1 is shown in Fig. 4.

Prediction / True	Positive	Negative
Positive	TP (7.2%)	FN (0.4%)
Negative	FP (2.2%)	TN (90.2%)

Fig. 4 Detection result of sensor1 for abnormal or not

Table 3 Classification results for 4 representative sensors

Accuracy	2 Missing	3 Minor	4 Outlier	5 Square	6 Trend	7 Drift
Sensor 1	0.99	0.95	0.96	0.99	0.96	0.97
Sensor 13	0.97	0.95	1	1	0.99	0.98
Sensor 25	1	1	1	0.99	1	1
Sensor 34	1	1	1	0.96	1	1

2.5 Data detection of various abnormalities

For simplicity, the result of four sensors which represent 4 different abnormal types respectively are shown in Table 3. For each sensor, the accuracy for each abnormal type reaches up to 95%, exhibiting well performance of the classification method. In the rest of the paper, features used to accurately distinguish each type will be fed into BiLSTM neural network.

2.6 Positioning of occurrence time

On the basis of two-level anomaly detection, the time scale of the algorithm is reduced to further precisely locate the time of anomaly occurrence. The time scale of various abnormality discrimination algorithms is used to cut down and iterate through abnormal data to realize the precise location of the abnormal occurrence time and extend the accuracy to the minute level. For further visualization, a box to mark the occurrence of abnormality, the text description, and comments to the relative information of the abnormality are added. The location and visualization of jump points and square abnormalities are as follows:

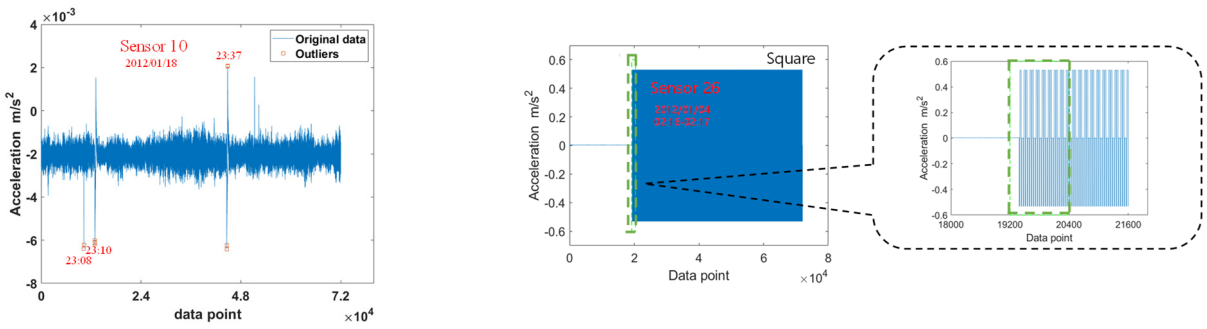
3. Anomaly prediction using BISTM

The basic principle of deep learning for data anomaly detection is to take advantage of the powerful classification function of BiLSTM based on a large amount of labeled data, and the network learns the features of the anomalous data automatically by supervised learning to achieve automatic classification. The benefit is that after building the network, the training data set of each sensor is input. After the training is completed, the model is saved and then the subsequent data can be automatically classified.

3.1 Deep learning

With the full swing of informatization and intelligence in various industries, data anomaly detection as a prerequisite for data cleaning is in high demand nowadays. Huge advances in artificial intelligence and deep learning offer new ways to automate the identification of data anomalies. For automated anomaly detection of data, we leverage the idea of LSTM Recurrent Neural Network. These types of networks excel at finding complex relationships in multivariate time series data. The basic idea of anomaly detection with LSTM neural network is: the system supervises previous values over hours of days and predicts the behavior of the next minute period (Hochreiter and Schmidhuber 1997). If the actual value a minute later is within, let's say, one standard deviation, it is normal data. If it is more, it is an anomaly. In this very study, the system could propagate data forward as well as backward in time, i.e., in bidirectional ways. Thus, a bidirectional LSTM neural network is finally settled to fulfill the task of data anomaly detection.

LSTM network architecture was originally developed by Hochreiter and Schmidhuber (1997). In detail, an input sequence vector $x = (x_1, x_2, \ldots, x_n)$ is given, where n indicates the length of the input sequence. The core structure of the LSTM is the usage of three gates to control a memory cell activation vector c. The first forget gate determines how much of the cell state c_{t-1} at the previous time is retained until the current cell state c_t; the second input gate determines the extent to which the input x_t of the network is saved to the current cell state c_t; the third output gate determines how much of the cell state c_t is transmitted to the current output value h_t of the LSTM networks. The three gates are a fully connected layer, and its input is a vector and the output is a real number in [0, 1]. The basic LSTM cell architecture is shown in Fig. 6, and its representation is as follows



(a) Positioning of outliers occurrence (b) Positioning of square occurrence

Fig. 5 Positioning of abnormal occurrence in minute-level

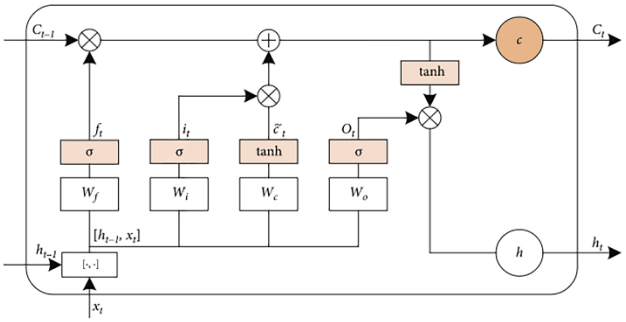


Fig. 6 Diagram of LSTM cell

$$\text{Input gates:} \quad i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{6}$$

$$\text{Output gates:} \quad o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{7}$$

$$\text{Forget gates:} \quad f_t = \sigma\left(W_{fx}x_t + W_{fh}h_{t-1} + b_f\right) \tag{8}$$

$$\text{Cell states:} \quad c_t = f_t * c_{t-1} + i_t * \tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c) \tag{9}$$

$$\text{Cell outputs:} \quad h_t = o_t \times \tanh(c_t) \tag{10}$$

where σ is the logistic sigmoid function, x_t indicates t-th vector of the sequence and h_t indicates the hidden state, W terms and b terms, respectively, represent weight matrices (e.g., W_{fx} represents the forget gate weight matrix) and bias vectors (e.g., b_i represents the input gate bias vector) for the three gates.

To overcome the shortcoming of the single LSTM cell that can only capture previous values but not utilize future values, Schuster and Paliwal invented bidirectional recurrent neural networks (BRNN) to combine two separate hidden LSTM layers of opposite directions to the same output (Schuster and Paliwal 1997). With this structure, the output layer is able to utilize related information from both the previous and future values. A BiLSTM model calculates the input sequence $x = (x_1, x_2, \ldots, x_n)$ from the opposite direction to a forward hidden sequence $\vec{h}_t = (\vec{h}_1, \vec{h}_2, \ldots, \vec{h}_n)$ and a backward hidden sequence $\overleftarrow{h}_t = (\overleftarrow{h}_1, \overleftarrow{h}_2, \ldots, \overleftarrow{h}_n)$. The encoded vector y_t is formed by the concatenation of the final forward and backward outputs, $y_t = [\vec{h}_t, \overleftarrow{h}_t]$.

$$\vec{h}_t = \sigma\left(W_{\vec{h}x}x_t + W_{\vec{h}\vec{h}}\vec{h}_{t-1} + b_{\vec{h}}\right), \tag{11}$$

$$\overleftarrow{h}_t = \sigma\left(W_{\overleftarrow{h}x}x_t + W_{\overleftarrow{h}\overleftarrow{h}}\overleftarrow{h}_{t+1} + b_{\overleftarrow{h}}\right), \tag{12}$$

$$y_t = W_{y\vec{h}}\vec{h}_t + W_{y\overleftarrow{h}}\overleftarrow{h}_t + b_y \tag{13}$$

where $y = (y_1, y_2, \ldots, y_t, \ldots, y_n)$ is the output sequence of the first hidden layer.

3.2 Feature extraction

Most deep learning algorithms can directly process time-series data, and feature extraction is not required to complete the classification work. However, the computation efficiency will decrease drastically if there is uncertainty in the data. Considering such an amount of data in this study, feature extraction is an inevitable option. For the convenience of training, each period of acceleration data will be represented by 9 manually chosen features, and each feature is one vector. In this study, 9 features have been chosen to reduce the computation burden, and the feature extraction is conducted every 1 minute. Thus, for a training sample ranging over one-hour values, the feature extracted is given in terms of a matrix of 9×60. All the features are listed as follows:

(1) mean value of the data;
(2) median value of the data;
(3) standard deviation of the data;
(4) root mean square level of the data;
(5) difference between the maximum and minimum of the data;
(6) maximum value divided by 80th percentile of the data;
(7) mean value divided by the amplitude of the data;
(8) mean value divided by 80th percentile of the data;
(9) Natural frequency estimated by Welch's power spectral density.

All these features are selected in the anomalous feature extraction process of the aforementioned manual empirical method and show high accuracy in the neural network classification task.

3.3 Experiment design

Hitherto, the input sequence has been settled, and the next work is to design a proper neural network. In this paper, two BiLSTM layers are stacked so that the model is capable of learning higher-level representations, followed by two fully connected layers. The architecture of the network is shown in Fig. 7. Input feature is a matrix of $n \times 9 \times 60$, where n represents the sample size. For simplicity, the number of neurons in each hidden layer is empirically set to the same of 16, the most popular

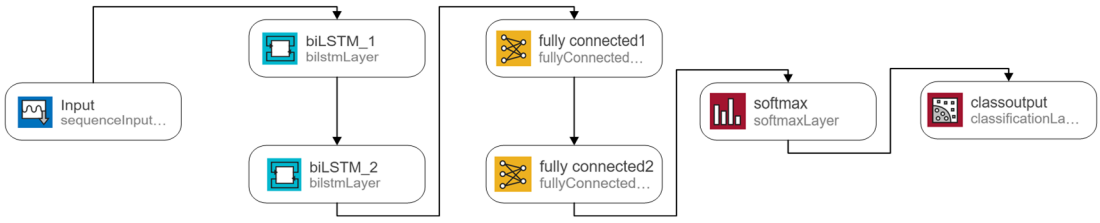


Fig. 7 Architecture of the proposed deep neural network

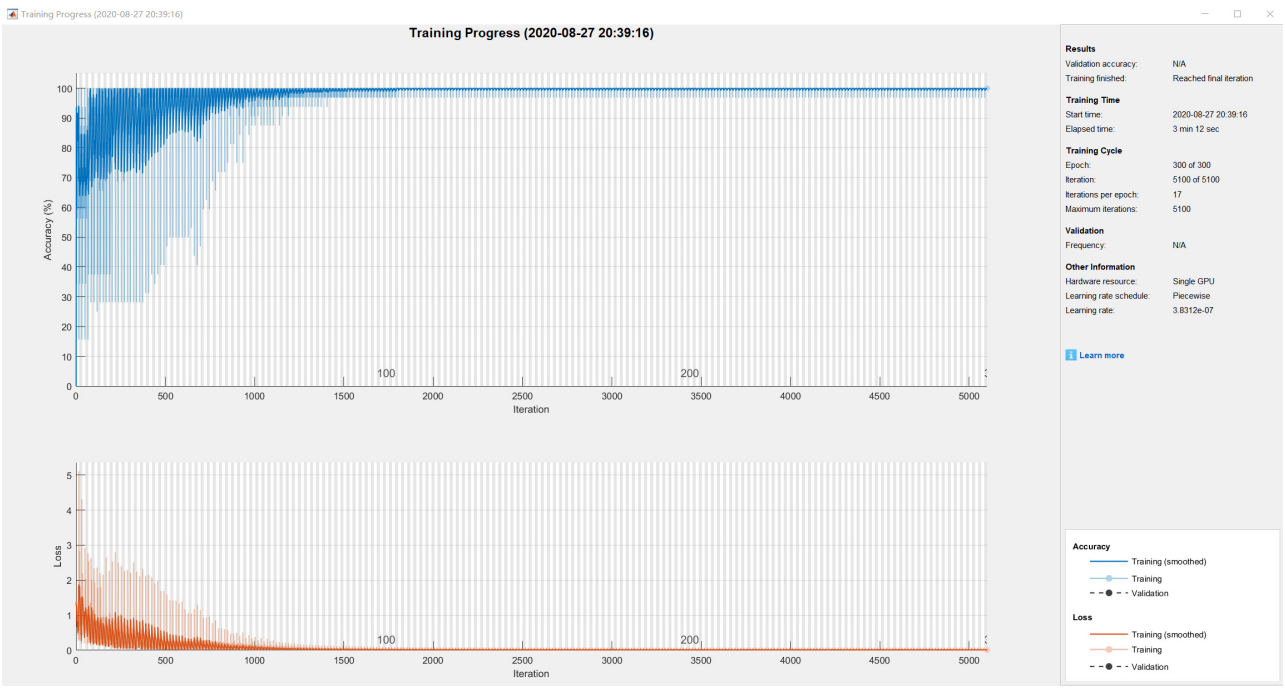
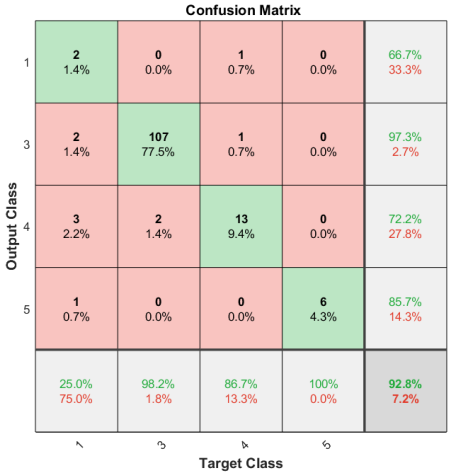
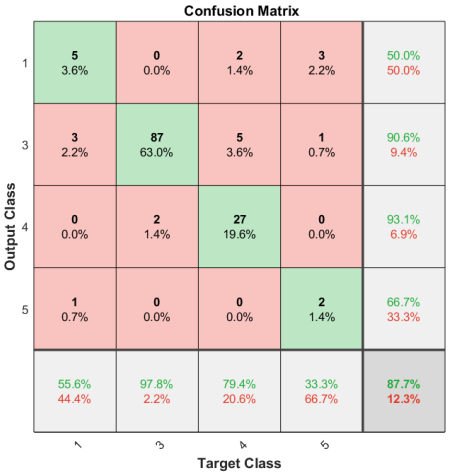


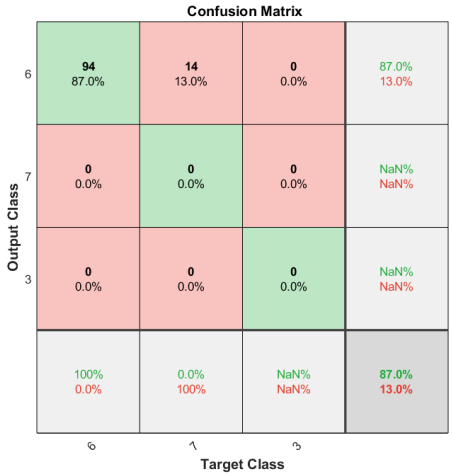
Fig. 8 Training progress



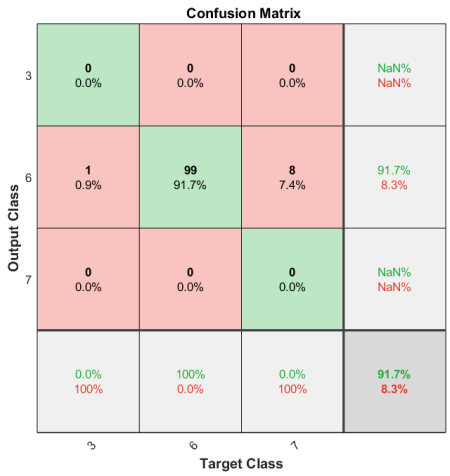
(a) Confusion matrix for sensor 1



(b) Confusion matrix for sensor 2

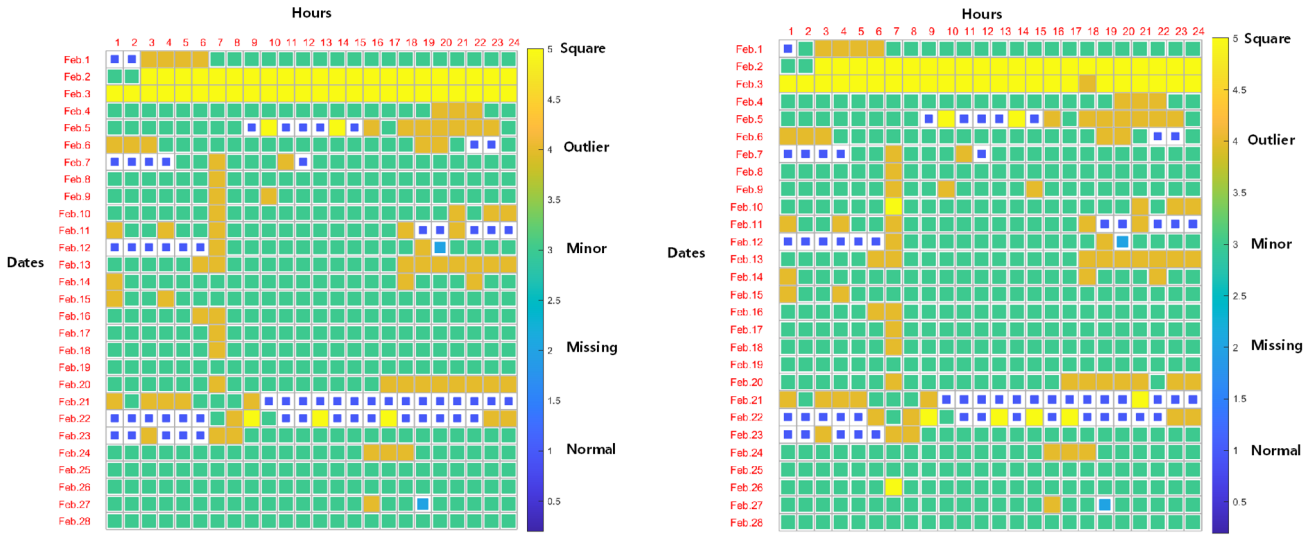


(c) Confusion matrix for sensor 13



(d) Confusion matrix for sensor 14

Fig. 9 Fault classification rates of different fault confusion matrix for proposed model



(a) Ground truth label (b) Predictive label

Fig. 10 Predicted result of sensor 1 per hours

Rectified Linear Unit (ReLu) is applied as the activation function of the hidden layers and Softmax is for the output layer. The learning rate was set as 0.01, and the optimizer is Adam. The maximum epoch is 300, each epoch takes about half minutes.

The proposed BiLSTM model is accomplished using MATLAB language. The workstation is configured with two Intel Xeon(R) E5-2696 v4 2.2 GHz CPUs, a 256 GB memory, and an NVIDIA TITAN X (Pascal) GPU.

The training progress is conducted by GPU acceleration. One archetypal example of training progress for one sensor is demonstrated in Fig. 8, where the accuracy and loss both converge gradually indicate the effective learning progress.

3.4 Anomaly prediction

The accuracy of the proposed BiLSTM neural network for each sensor is depicted in Fig. 9 (some examples are shown for simplicity). In the confusion matrix, the target class and the output class denote the true class and the model classification prediction class, respectively. The number above each square in the confusion matrix indicates the number of samples that output the true category as the predicted category and their corresponding percentages. Take sensor 1 as an example (shown in Fig. 9(a)), the first five diagonal cells show the number of percentages of correct classification by the trained network. 107 type 2 faults are correctly classified to the corresponding types, 13 cases are correctly classified as type 3 faults, 2 cases of type 1 faults are classified as type 2 faults, and 1 case of type 4 fault is classified as type 2 fault. Generally, the accuracy of correctly classified as type 2 is up to 97.3%, and the overall accuracy of sensor 1 is up to 92.8%. Similarly, the overall classification accuracy for sensor 2 is 87.7%, 87.0% for sensor 13, and 91.7% for sensor 14, which displays the good performance of the proposed BiLSTM neural network model. Overall, the classification accuracy of all sensors is above 87%, which demonstrates the reliability and accuracy of the BiLSTM deep learning method. It is worth pointing out that the missing type is not identified in the deep learning approach, because the missing data directly prevents the computation from being performed. The missing type has been identified in the first step by traditional methods, and the recognition rate for this type is 100%.

3.5 Test set validation

Using the supplemented February data as the test set to further validate the deep learning model. The accuracy of the model on the new test set was tested. Taking sensor 1 as an example to illustrate the generalization capability of the model. As shown in Fig. 10, 85%, 97%, 97%, and 100% accuracy were achieved for anomaly types 1, 3, 4, and 5, respectively. Missing types are identified by traditional methods with 100% accuracy. The accuracy of each type of anomaly identification for other sensors is also above 85%. This shows the high accuracy and stability of the method in this paper on the test sets.

4. Conclusions

4.1 Anomaly prediction

The overall principles and ideas of abnormal prediction include:

- The false negatives error type should be strictly controlled. Because mistaking abnormal data as normal data will bring great difficulties to subsequent data processing and leading to misjudgment of structural state.
- Missing is a special anomaly, owing to the remarkable performance it can be distinguished easily and directly. While it leads to difficulty in calculating for intelligent algorithms. Therefore, when using intelligent algorithms to make predictions, it is recommended to use traditional

Table 4 Comparison of the two kinds of methods

Method	Fundamental principle	Crucial point	Accuracy	Feasibility
Artificial experience method	Time-frequency domain statistical characteristics	Proper parameter setting by manual tuning	95-100%	Higher accuracy rate owing to the fine-tuned thresholds for each sensor while leads to labor-intensive.
BiLSTM Deep learning method	Supervised learning by neural network	Good feature input and reasonable network architecture	87-93%	A high degree of automation with an acceptable accuracy rate.

manual methods to eliminate missing data before attempting intelligent algorithms.

4.2 Feature extraction

The selection of time scale in feature extraction is worthy of serious consideration. The final time scale for identification is a classification label per hour. If the feature extraction of the original data is performed at this scale, the overall trend characteristics of many data will be masked. However, if the time scale is too small, it will greatly increase the amount of calculation and cause data redundancy. The dimensionality can be different for different features, but it is recommended that the sampling frequency be no less than the one-minute scale. When selecting specific features, both time-domain features and frequency-domain features should be taken into consideration. For the feature extraction stage of deep learning, in addition to common statistical indicators such as mean-variance, it is best to combine the beneficial features obtained from previous manual experiences such as the proportion of data in the bandwidth and other feature parameters to combine the advantages of the two methods. As for the impact of feature selection, features are directly related to whether anomalies can be recognized for the artificial empirical method, and for the BiLSTM deep modeling method, in addition to affecting the accuracy of recognition also has different degrees of impact on the convergence speed and computation time.

4.3 Comparison of two types of methods

The artificial experience method based on various abnormal statistical characteristics has clear principles and is more accurate. Also, thanks to the clear logic of this method, various types of abnormal data can be treated in different ways. However, the artificial experience method requires a lot of manual intervention in adjusting the threshold of various indicators for each sensor, as a result, the feasibility performs poorly. As for the time-consuming calculation, the manual experience method is faster, and the average recognition time per hour is about less than 2 s. The calculation time of deep learning is mainly in model training, which takes about 3 hours. If just for prediction, the calculation time of hourly data is about 10 s.

As for the deep learning method, all the tedious calculations are completed by the computer, which is more intelligent and efficient. However, as the deep learning method is essentially a computational fuzzy learning method, it is difficult to implement rationalized manual intervention. The deep learning method has a greater probability of misjudgment for normal data compared to abnormal data, as a result, the accuracy of this method is not as high as that of the manual empirical method. The detailed comparison of the two methods is shown in Table 4.

Therefore, based on artificial experience, optimizing the feature vector is an effective way to combine artificial experience and deep learning. In this way, the advantages of the two methods can be combined to achieve a satisfactory accuracy anomaly detection rate and reduce the tedious manual threshold parameter setup as much as possible.

Acknowledgments

The authors would like to thank the organizations of the International Project Competition for SHM (IPC-SHM 2020) ANCRiSST, Harbin Institute of Technology (China), and University of Illinois at Urbana-Champaign (USA) for their generously providing the invaluable data from actual structures. The authors also would like to thank the chairs of IPC-SHM 2020 Prof. Hui Li, and Prof. Billie F. Spencer Jr. for their leadership on the competition.

Also, the authors sincerely acknowledge financial support from the National Natural Science Foundation of China (Grants. 51978154), the Fund for Distinguished Young Scientists of Jiangsu Province (Grant. BK20190013), Key Research and Development Program of Nanjing Jiangbei New Area (Grant. ZDYF20200118) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX21_0113).

References

Bao, Y. and Li, H. (2020), "Machine learning paradigm for structural health monitoring", *Struct. Health Monitor.*, **20**(4), 1353-1372. https://doi.org/10.1177/1475921720972416

Bao, Y., Chen Z., Wei, S., Xu, Y., Tang, Z. and Li, H. (2019a), "The state of the art of data science and engineering in structural health monitoring", *Engineering*, **5**(2), 234-242. https://doi.org/10.1016/j.eng.2018.11.027

Bao, Y., Tang, Z., Li, H. and Zhang, Y. (2019b), "Computer vision and deep learning-based data anomaly detection method for structural health monitoring", *Struct. Health Monitor.*, **18**(2), 401-421. https://doi.org/10.1177/1475921718757405

Chen, Z., Bao, Y., Li, H. and Spencer, B. (2018a), "A novel distribution regression approach for data loss compensation in structural health monitoring", *Struct. Health Monitor.*, **17**(6), 1473-1490. https://doi.org/10.1177/1475921717745719

Chen, Z., Li, H. and Bao, Y. (2018b), "Analyzing and modeling

inter-sensor relationships for strain monitoring data and missing data imputation: a copula and functional data-analytic approach", *Struct*. *Health Monitor*., **18**(4), 1168-1188. https://doi.org/10.1177/1475921718788703

Chen, Z., Bao, Y., Li, H. and Spencer, B. (2019), "LQD-RKHS-based distribution-to-distribution regression methodology for restoring the probability distributions of missing SHM data", *Mech*. *Syst*. *Signal Process*., **121**, 655-674. https://doi.org/10.1016/j.ymssp.2018.11.052

Hochreiter, S. and Schmidhuber, J. (1997), "Long short-term memory", *Neural Comput*., **9**(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, Y., Beck, J.L., Wu, S. and Li, H. (2016), "Bayesian compressive sensing for approximately sparse signals and application to structural health monitoring signals for data loss recovery", *Probabil*. *Eng*. *Mech*., **46**, 62-79. https://doi.org/10.1016/j.probengmech.2016.08.001

Li, L., Liu, G., Zhang, L. and Li, Q. (2019), "Sensor fault detection with generalized likelihood ratio and correlation coefficient for bridge SHM", *J*. *Sound Vib*., **442**, 445-458. https://doi.org/10.1016/j.jsv.2018.10.062

Liu, G., Li, L., Zhang, L., Li, Q. and Law, S.S. (2020), "Sensor faults classification for SHM systems using deep learning-based method with Tsfresh features", *Smart Mater*. *Struct*., **29**(7), 075005. https://iopscience.iop.org/article/10.1088/1361-665X/ab85a6

Liu, H., Shah, S. and Jiang, W. (2004), "On-line outlier detection and data cleaning", *Comput*. *Chem*. *Eng*., **28**(9), 1635-1647. https://doi.org/10.1016/j.compchemeng.2004.01.009

Kullaa, J. (2013), "Detection, identification, and quantification of sensor fault in a sensor network", *Mech*. *Syst*. *Signal Process*., **40**(1), 208-221. https://doi.org/10.1016/j.ymssp.2013.05.007

Ni, F., Zhang, J. and Noori, M.N. (2019), "Deep learning for data anomaly detection and data compression of a long-span suspension bridge", *Comput*.*-Aided Civil Inf*., **35**(7), 685-700. https://doi.org/10.1111/mice.12528

Schuster, M. and Paliwal, K.K. (1997), "Bidirectional recurrent neural networks", *IEEE T Signal Process*., **45**(11), 2673-2681. https://ieeexplore.ieee.org/document/650093

Spencer Jr., B.F., Hoskere, V. and Narazaki, Y. (2019), "Advances in computer vision-based civil infrastructure inspection and monitoring", *Engineering*, **5**(2), 199-222. https://doi.org/10.1016/j.eng.2018.11.030

Tang, Z., Chen, Z., Bao, Y. and Li, H. (2019), "Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring", *Struct*. *Control Health*., **26**(1), e2296. https://onlinelibrary.wiley.com/doi/abs/10.1002/stc.2296

Xia, Y. and Ni, Y. (2018), "A wavelet-based despiking algorithm for large data of structural health monitoring", *Int*. *J*. *Distrib*. *Sens*. *N*, **14**(12),1550147718819095. https://doi.org/10.1177/1550147718819095

Yang, Y. and Nagarajaiah, S. (2016), "Harnessing data structure for recovery of randomly missing structural vibration responses time history: sparse representation versus low-rank structure", *Mech*. *Syst*. *Signal Process*., **74**, 165-182. https://doi.org/10.1016/j.ymssp.2015.11.009

Yi, T., Huang, H. and Li, H. (2017), "Development of sensor validation methodologies for structural health monitoring: a comprehensive review", *Measurement*, **109**, 200-214. https://doi.org/10.1016/j.measurement.2017.05.064

Yu, M., Wang, D. and Luo, M. (2013), "Model-based prognosis for hybrid systems with mode-dependent degradation behaviors", *IEEE Trans*. *Ind*. *Electron*., **61**(1), 546-554. https://ieeexplore.ieee.org/document/6425465

Yuen, K.-V. and Ortiz, G.A. (2017), "Outlier detection and robust regression for correlated data", *Comput*. *Method Appl*. *Mech*. *Eng*., **313**, 632-646. https://doi.org/10.1016/j.cma.2016.10.004