

## Research Article

# Structural Vibration Data Anomaly Detection Based on Multiple Feature Information Using CNN-LSTM Model

Xiulin Zhang <sup>1,2,3</sup> and Wensong Zhou <sup>1,2,3</sup>

<sup>1</sup>Key Lab of Smart Prevention and Mitigation of Civil Engineering Disasters of the Ministry of Industry and Information Technology, Harbin Institute of Technology, Harbin 150090, China

<sup>2</sup>Key Lab of Structures Dynamic Behavior and Control of the Ministry of Education, Harbin Institute of Technology, Harbin 150090, China

<sup>3</sup>School of Civil Engineering, Harbin Institute of Technology, Harbin 150090, China

Correspondence should be addressed to Wensong Zhou; [zhouwensong@hit.edu.cn](mailto:zhouwensong@hit.edu.cn)

Received 11 March 2023; Revised 30 June 2023; Accepted 27 July 2023; Published 5 August 2023

Academic Editor: Francesc Pozo

Copyright © 2023 Xiulin Zhang and Wensong Zhou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Structural health monitoring (SHM) system has been operating for a long time in a harsh environment, resulting in various abnormalities in the collected structural vibration monitoring data. Detecting these abnormal data not only requires user interaction but also is quite time-consuming. Inspired by the manual recognition process, a vibration data anomaly detection method based on the combined model of convolutional neural network (CNN) and long short-term memory (LSTM) network is proposed in this paper. This method simulates intelligent human decision making in two steps. First, the original data are reconstructed by two feature sequences with higher universality and smaller size. In the time domain, the residual signal is extracted from the upper and lower peak envelopes of the original data to characterize the symmetry of the data. In the frequency domain, the power spectral density sequence of the original data is extracted to characterize the interpretability of the data. Second, a CNN-LSTM model is constructed and trained which utilizes CNN to extract local high-level features of input sequence and inputs new continuous high-level feature representations into LSTM to learn global long-term dependencies of abnormal data features. For verification, the method was applied to the automatic classification of continuous monitoring data for 42 days of long-span bridge, and the average accuracy of the classification results exceeded 94% and the detection time was 78 minutes. Compared with existing methods, this method can detect abnormal data more accurately and efficiently and has a stronger generalization ability.

## 1. Introduction

Vibration-based structural health monitoring (SHM) [1] methods involve tasks such as structural modal analysis [2], model updating [3, 4], damage detection [5–7], and safety assessment [8, 9]. The analysis results rely heavily on having accurate and high-quality vibration data. However, in real-time and long-term SHM systems, the vibration data collected automatically might be abnormal owing to sensor faults and the transmission or storage failures caused by the harsh environment [10], especially after long periods of service. Anomalous data will almost certainly result in incorrect identification results and false warnings. In addition,

when these anomalies are mixed with data related to emergency events, such as earthquakes, ship collisions, or traffic accidents, they will also interfere with the early warning capability of the SHM system. Therefore, identifying and locating abnormal data is an essential preprocessing step for vibration data analysis.

Data anomaly detection, based on physical model construction and probabilistic prediction, is a common and effective strategy. For example, Thiyagarajan et al. [11] detected the abnormality of a sensor based on the autoregressive integrated moving average (ARIMA) model and provided a fault warning. Li et al. [12] proposed a rapid sensor fault identification method based on the generalized

likelihood ratio and correlation coefficient. Wan and Ni [13] employed a Bayesian modeling method using Gaussian processes (GPs) to detect abnormal data via probabilistic prediction of the structural stress response. However, as system complexity and uncertainty increase, especially when massive and continuous monitoring data exist, it is challenging to find explicit models with appropriate parameters that remain computationally efficient [14].

Considering that the continuous monitoring of structures can produce massive amounts of data, data-driven methods, which do not require a physical model of the system, have become more popular in recent years. For example, deep learning methods such as CNN models and LSTM models have been widely used in the field of SHM [15–28]. These methods have the greatest potential to learn from monitoring data containing abnormal data and to automatically diagnose various abnormal patterns. According to the algorithm features, basic model, and input data type, the corresponding methods can be divided into three categories.

The first includes computer vision methods, which take advantage of visualizing monitoring data. For example, Bao et al. [15] converted time-series signals into images, inputted them into a deep neural network (DNN), and then trained the model to detect anomalies. Tang et al. [16] fused the frequency-domain features, constructed a dual-channel image, and incorporated a CNN to complete the identification and classification of abnormal acceleration data. Mao et al. [17] converted time-series data into a Gramian angular field image and identified abnormal data using a combination of a generative adversarial network (GAN) and autoencoders. However, segmenting and converting signals into images to extract features is time-consuming, and it is difficult to process continuous long-term monitoring data. Most importantly, it is easy to lose key information during the visualization process.

The second type uses a time-series prediction model to detect abnormal data by observing the difference between the actual and predicted values. For example, Zhang et al. [18] used historical normal data to train a separate LSTM network for each sensor and then set a threshold for the prediction error of the network to detect anomalies. Vos et al. [19] combined LSTM with a class of support vector machines and separated the abnormal data collected during the durability test of the reduction gearbox from the normal vibration signal according to the residual signal between the actual and predicted values. This type of method is usually limited because it works only on specific sensors and is difficult to extend to other datasets. Furthermore, it cannot be used for more detailed classification of multiple anomalies. In addition, it is challenging to directly use the original data to generate forecast data in terms of calculation cost.

The third employs a time-series classification model to detect abnormal data. Yang et al. [20] constructed a new time series by extracting nine feature indexes from the original data and combined this result with a bidirectional long short-term memory (Bi-LSTM) neural network model to classify and locate GPS data anomalies. Zhang and Lei [21]

extracted the maximum and minimum values of acceleration data by downsampling to reduce the dimension of input samples and classified anomalous data in combination with a 1D-CNN. For confusing patterns, Zhang et al. [22] used statistical features to reclassify the intermediate results of CNN model recognition. Based on the shape of the original time series, Arul and Kareem [23] combined the shapelet transform with a random forest classifier to detect anomalies in SHM data. Although these approaches are effective, some problems still need to be addressed. The resampled feature signal lacks the frequency-domain information of the original signal. In addition, when quantitative feature indicators, such as the maximum and mean values, are used as training samples, the model will only be effective on datasets with specific structures, which means that the generalization ability of the model is low.

To address the above challenges, there remains a need for an efficient method with a strong generalization ability to automatically detect abnormal data. Inspired by existing methods, this study proposes a vibration data anomaly detection method based on CNN-LSTM models. This method reconstructs and marks original data by extracting multiple feature information from the time and frequency domains, and the reconstructed samples have more obvious generality and smaller size, which will improve the classification efficiency of the classifier. This method combines the advantages of both CNN and LSTM models, utilizing CNN to extract local high-level features of input samples and inputting new continuous high-level feature representations into LSTM to learn global long-term dependencies. Therefore, the combined model can more accurately learn abnormal features in vibration data. The example results for long-span bridges show that the proposed method has a higher efficiency, accuracy, and generalization ability compared to previous methods.

The remainder of this paper is organized as follows. Section 2 describes the framework of the proposed method in detail, including the feature sequence extraction method and the working principle of the CNN-LSTM model. Section 3 presents a detailed example of a long-span suspension bridge. The results show that the proposed method is both efficient and accurate. Section 4 provides further discussion. Finally, concluding remarks are given in Section 5.

## 2. Methodology

An overview of the proposed data anomaly detection method based on the CNN-LSTM model is shown in Figure 1. First, the proposed feature sequences are extracted from the time and frequency domains of the segmented data. The size of the original monitored data is significantly reduced, which helps to improve the classification accuracy and efficiency. Second, a CNN-LSTM model that applies time-series analysis is constructed and trained for anomaly data classification. Finally, with well-trained models, these potential anomalies can be automatically detected in the test set composed of actual structural vibration data, which can replace manpower in large-scale data detection.

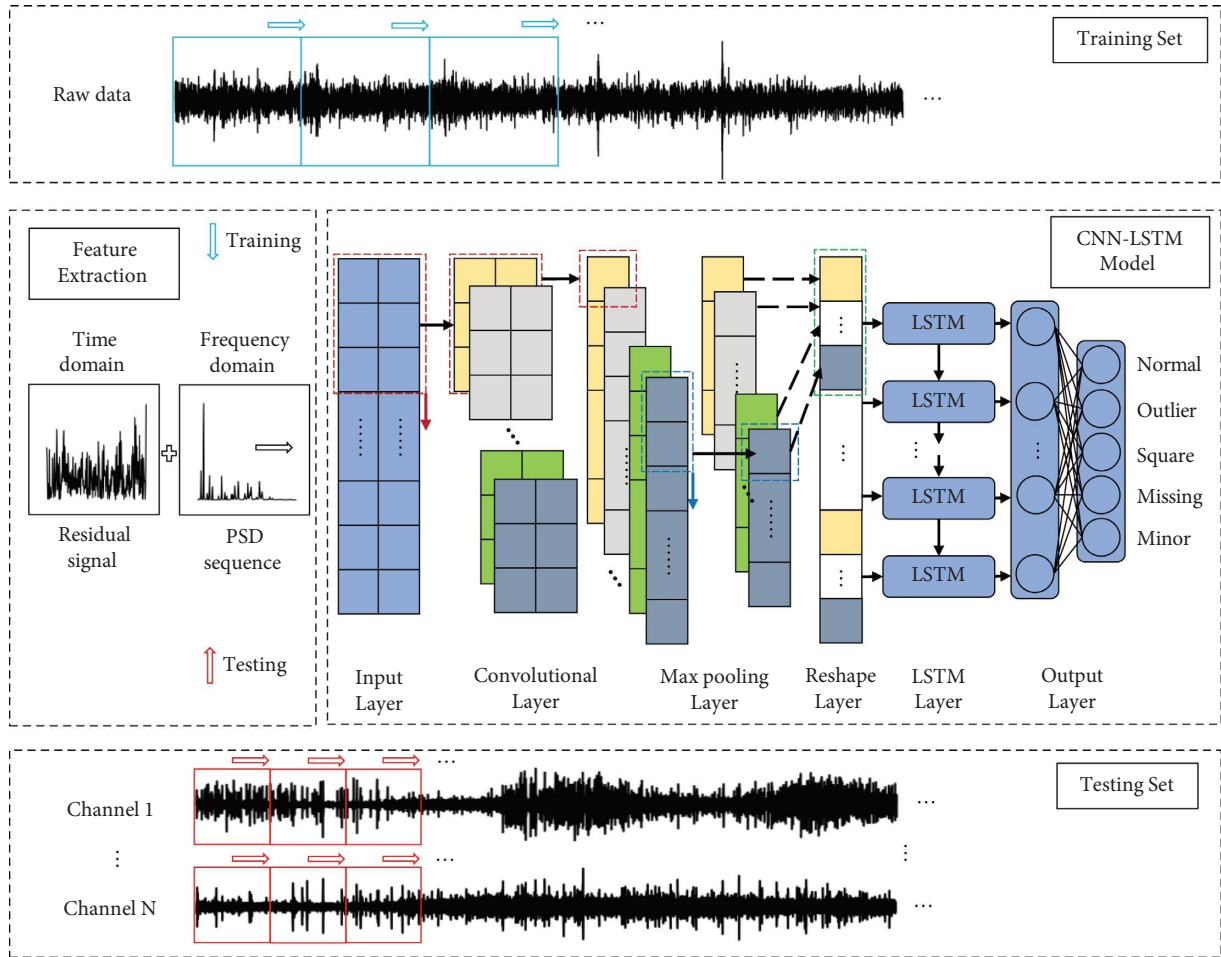


FIGURE 1: Framework of the proposed method.

**2.1. Data Pattern Classification.** Vibration monitoring data can be divided into seven categories based on local characteristics [15]: normal, trend, outlier, drift, square, missing, and minor. As shown in Figure 2, typical vibration signals with only a single feature are presented. Figure 2(a) shows the normal pattern data as a reference. Figures 2(b)–(d) show the outlier pattern, trend pattern, and drift pattern of abnormal vibration data. In this study, we found that the number of samples of pure trend pattern and drift pattern caused by sensor failure is very small, and they usually appear in the same sample mixed with outlier pattern. As shown in Figure 3, it is difficult to accurately define this mixed data pattern, and this mixed pattern is widespread and numerous. Considering that these three types exhibit strong asymmetry in the time domain and poor interpretability in the frequency domain compared to normal data, this study categorizes them into one category: outlier pattern. Figures 2(e) and 2(f) show two abnormal patterns: square and missing, of which the data of the former are like a square wave, and the latter is completely or partially missing. Figure 2(g) shows the abnormal data of the minor pattern and the normal monitoring data collected by adjacent sensors during the same period. Compared to the normal data, the amplitude of the minor pattern was very small in the time domain. Therefore, the dataset used in this

study divides vibration monitoring data into five categories: normal, outlier, square, missing, and minor. In addition, when a sample has multiple anomaly features, it corresponds to only one real label. This requires prioritizing each type of vibration data for a single classification. In this study, the missing pattern had the highest priority, followed by outlier, square, minor, and normal with the lowest priority.

**2.2. Feature Sequence in Time Domain.** In this study, the feature sequence of the vibration data in the time and frequency domains was extracted as reference indexes for classification.

In the time domain, the feature sequence is obtained from the profile of the vibration signals, which is approximately symmetric about the time axis. The structural vibration is generally an up-and-down reciprocating motion around the equilibrium position, and the absolute values of the up-and-down amplitudes are approximately equal. For the vibration data under environmental excitation, the upper and lower peaks are approximately symmetric with respect to the time axis. For the vibration data under the heavy vehicle load, there is usually a higher peak followed by a slightly lower peak, and the vibration data are also symmetrical. This symmetry can be expressed using two

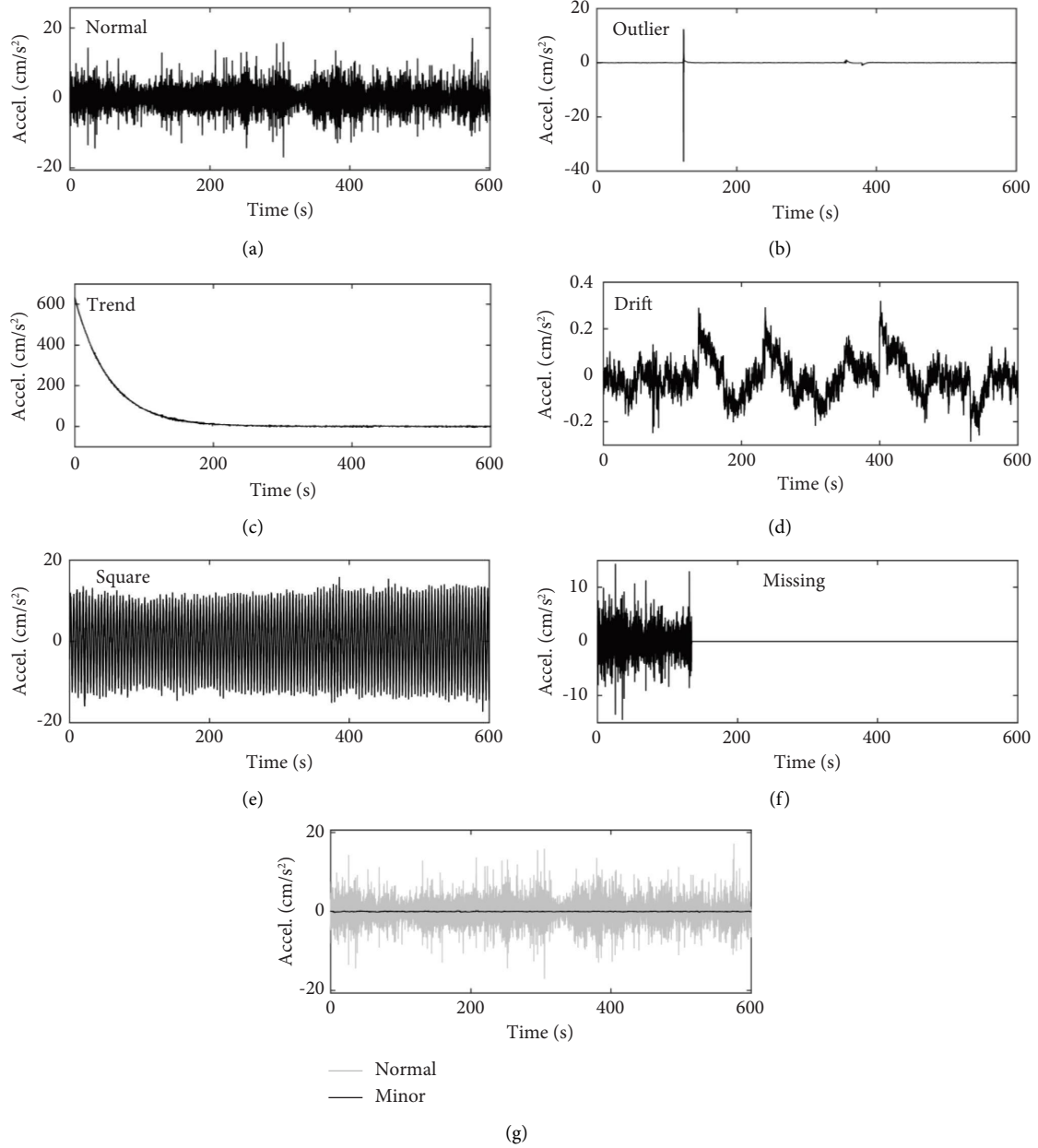


FIGURE 2: Visualization of various types of vibration data: (a) normal; (b) outlier; (c) trend; (d) drift; (e) square; (f) missing; (g) minor.

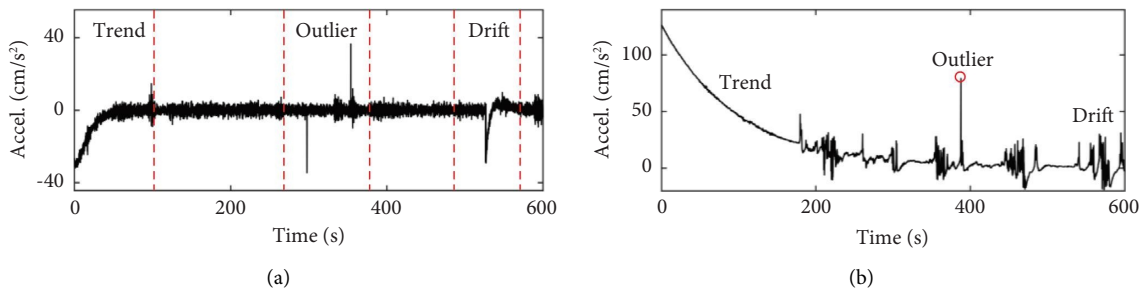


FIGURE 3: Visualization of widely distributed mixed pattern data.

envelope curves, as shown in Figure 4. In the figure, the gray curve represents the original data, the red curve represents the upper peak envelope, and the blue curve represents the lower peak envelope. Amplitude symmetry manifests as the symmetry of the envelope curves. However, this symmetry is not strictly a point-to-point correspondence in sequence; for example, point 1 in the figure corresponds to point 2, and point 3 corresponds to point 4. The reason for this is the existence of a vibration phase difference.

Taking the outlier as an example, when there are obvious anomalies in the data, the symmetry is broken, as indicated by the light blue dotted line box in Figure 5(b), and the absolute values of the upper and lower peak envelopes differ considerably. Therefore, we can estimate the symmetry of  $n$ -dimensional vibration data to detect abnormal patterns. The specific method is to extract the residual signal between the upper peak envelope  $\mathbf{y}_u \in \mathbb{R}^{n \times 1}$  and lower peak envelope  $\mathbf{y}_l \in \mathbb{R}^{n \times 1}$  of the original signal to reflect this symmetry. The original signal is considered symmetrical about the zero axis, and the residual signal is generated using the following formula:

$$\begin{cases} \mathbf{r} = [r_1, r_2, \dots, r_i], \\ r_i = \frac{|y_u^i + y_l^i|}{|x_u^i - x_l^i|}, \end{cases} \quad (1)$$

where  $i$  is the serial number of the sampling window, and when the window size is  $m$ ,  $n = i \times m$  is satisfied.  $y_u^i$  and  $y_l^i$  are the maximum value of the upper peak envelope and the minimum value of the lower peak envelope in window  $i$ , respectively, as shown in Figure 5.  $x_u^i$  and  $x_l^i$  are the maximum value of the upper peak envelope and the minimum value of the lower peak envelope, respectively, on window  $i$  after removing outliers from the original data, as shown in Figure 6. Here, the outliers are defined as elements that differ from the local average value by more than six times the local standard deviation within the specified window  $i$ . The element at the location of the outlier is replaced by the local average value and the peak envelope is recalculated. The purpose is to change the residual signal into a signal composed of dimensionless relative values, and the relative value of the residual is not significantly reduced owing to the removal of local outliers. In addition, for missing data, all "NaN" and "INF" values are replaced with zero to avoid calculating invalid values using equation (1). Equation (1) shows that when  $y_u^i$  and  $y_l^i$  calculated in window  $i$  have the same sign or when the two have different signs but their absolute values differ greatly, symmetry will be lost, and the calculated residual value will be greater than 1. When the two have different signs but their absolute values are approximately equal, the residual value will be very small and less than 1.

The peak envelope of each data pattern is shown on the left side of Figure 7, and the middle shows a comparison between the upper and lower peak (taking the absolute value) envelopes of each data pattern. When the data exhibit a normal pattern, the fitting results of the two curves are

better. In contrast, when outlier values exist in the data, the fitting effect is very poor. The fitting results are the best for the square pattern. For the missing pattern, the values of the two curves are zero. When the data have a minor pattern, this fitting effect worsens. The residual signal of each data pattern calculated using equation (1) is shown on the right side of Figure 7. Most residual values of the normal samples are below 1. For the residual signal of the outlier pattern, some values are much greater than 1, showing multiple peaks, while the value of the square residual is far less than 1, the value of the minor residual is slightly greater than 1, and the value of the missing residual signal is always zero. Based on these differences, we can take the residual signal as the feature sequence of the vibration data in the time domain. It has the advantage of having a much smaller size than the original data. The size of the feature sequence can be reduced by adjusting the size of window  $i$ ; as shown, we take the size of each window as  $m$ , thereby reducing the size of the feature sequence from  $n$  to  $n/m$ .

**2.3. Feature Sequence in Frequency Domain.** In the frequency domain, the feature sequence is obtained from the spectrum curves directly. The normal vibration data collected by SHM can reflect the dynamic characteristics of the structure, which is highly interpretable in the frequency domain and shows clear multi-peak characteristics, as shown in Figure 8(a). However, for various anomaly patterns, the characteristics of their power spectral density (PSD) curves are significantly different from those of normal data. For the outlier data, the PSD curve has no obvious peak, and the energy is concentrated near zero, as shown in Figure 8(b). For the square data, the PSD curve exhibits an obvious first-order peak, as shown in Figure 8(c). When the data are completely missing, the PSD curve is 0. When the data are partially missing, the form of the PSD curve depends on the type of data that are not missing, as shown in Figure 8(d). The minor pattern is similar to that of the outlier pattern, and the PSD curve has no obvious peak value and thus cannot be analyzed. Therefore, the PSD sequence is taken as the feature of the vibration data for anomaly detection. To ensure consistency with the length of the feature sequence in the time domain, the first  $i$  values of the PSD are used to construct the feature sequence and normalize each sample. The Fourier transform number can be adjusted to ensure that the feature sequence is within the frequency range of interest of the structure.

**2.4. CNN-LSTM Network Model.** The CNN-LSTM model structure [29] is mainly composed of input, convolutional, max pooling, reshape, LSTM, and output layers, as shown in Figure 9. In this study, the input layer in the figure consists of two feature sequences extracted from the original vibration samples. Each feature sequence is a one-dimensional time series; therefore, the model input is a  $2 \times i$  matrix. The feature series data are fed into the CNN convolution layer for convolution calculation, and high-level sequences of data features are extracted. The extracted feature matrix is pooled by the max pooling layer. The feature map is flattened to fuse

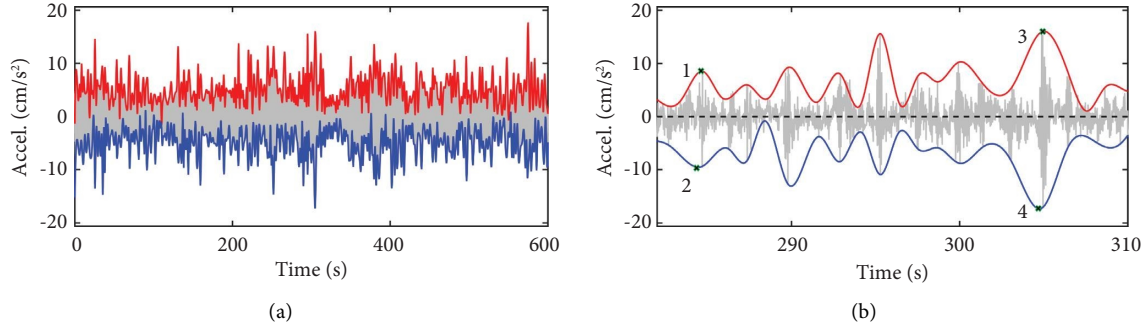


FIGURE 4: Symmetry of peak envelope of normal pattern data: (a) global view; (b) detail view.

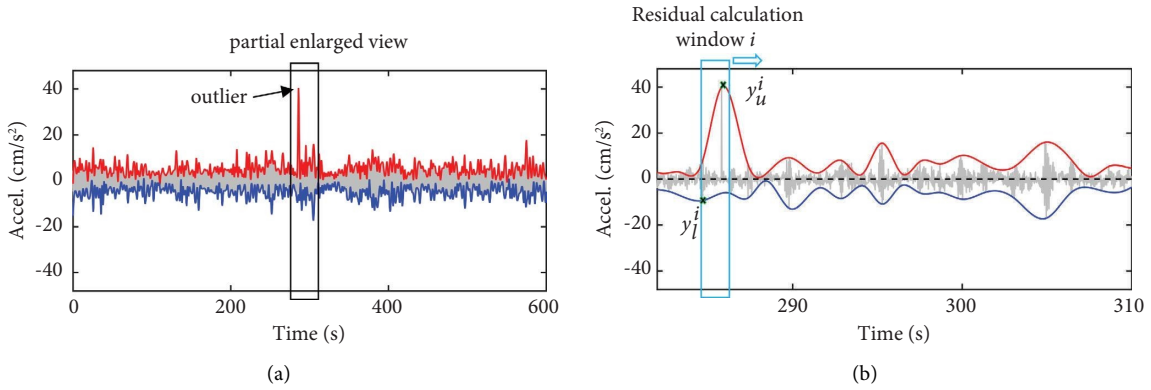


FIGURE 5: The peak envelope of outlier pattern data: (a) global view; (b) detail view.

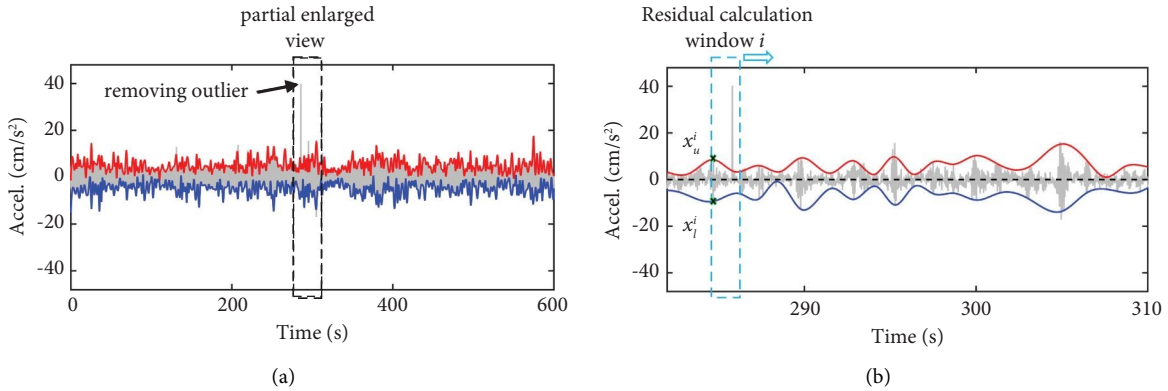


FIGURE 6: The peak envelope after removing outliers: (a) global view; (b) detail view.

the time- and frequency-domain features of the data. Then, input the flattened vector to the LSTM layer to capture the long-term dependence of the window feature sequence. Finally, the output of the LSTM layer is connected to the fully connected layer, and the softmax layer is used to classify the vibration anomaly characteristics to complete the anomaly detection process.

**2.4.1. Convolutional Neural Network.** The convolutional neural network constructed in this study uses one-dimensional convolution, which involves a filter vector

$\mathbf{m} \in \mathbb{R}^{k \times d}$  sliding over a sequence  $\mathbf{x} \in \mathbb{R}^{L \times d}$  and detecting features at different positions  $j$ . In Figure 9, the red dotted box represents the detected area, and the direction of the red arrow represents the sliding direction of the filter vector. For each position  $j$  in the sequence, we have a window vector  $w_j$  with  $k$  consecutive value vectors, denoted as

$$\mathbf{w}_j = [\mathbf{x}_j, \mathbf{x}_{j+1}, \dots, \mathbf{x}_{j+k-1}], \quad (2)$$

where,  $\mathbf{x}_j \in \mathbb{R}^d$  denotes the  $d$ -dimensional vector for the  $j$ -th value in the input sequence. A filter  $m$  convolves with the window vectors at each position in a valid manner to



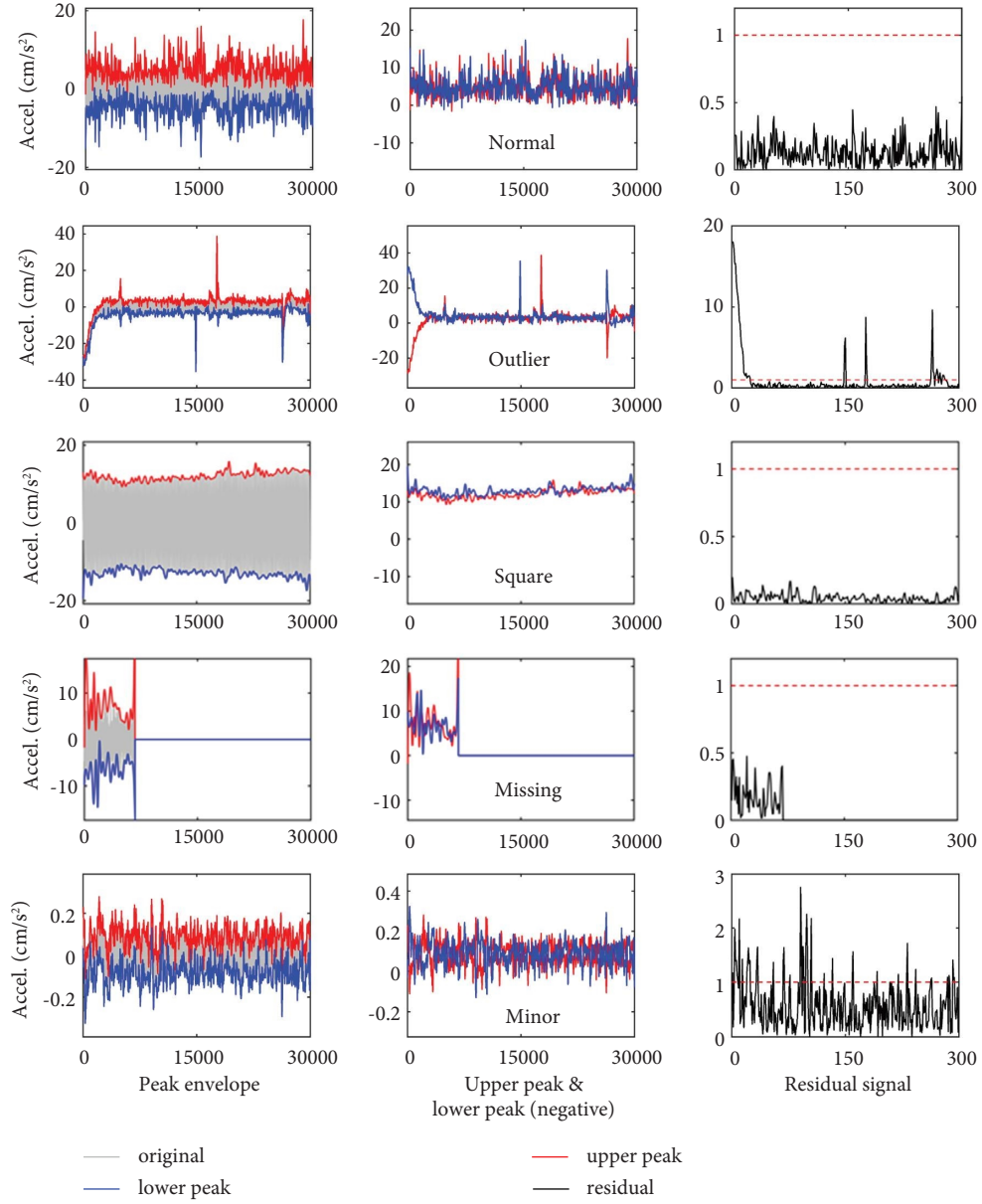


FIGURE 7: The residual signal of various data patterns.

generate a feature map  $\mathbf{e} \in \mathbb{R}^{L-k+1}$ , where each element  $e_j$  of the feature map for window vector  $w_j$  is produced as follows:

$$e_j = f(\mathbf{w}_j \otimes \mathbf{m} + b), \quad (3)$$

where  $\otimes$  denotes element-wise multiplication,  $b \in \mathbb{R}$  is a bias term, and  $f$  is a nonlinear transformation function. In this study, we selected the ReLU function [30].

The pooling layer is set after the convolution layer to continue reducing the dimension of the feature map  $e_i$  and improve the operation speed. For the pooling operation, max pooling is selected. The pooling window scans  $e_i$  from top to bottom, as shown in the blue dotted box in Figure 9, selects the maximum value in the pooling window as the output of this position, and finally obtains a smaller feature map  $c_i$ .

The CNN-LSTM model uses multiple filters to generate multiple feature maps. For  $n$  filters with the same length, the generated  $n$  feature maps can be rearranged as feature representations for each window  $w_j$ .

$$\mathbf{C} = [\mathbf{c}_1; \mathbf{c}_2; \dots; \mathbf{c}_n]. \quad (4)$$

Here,  $c_i$  is the feature map generated with the  $i$ -th filter and pooling and “;” represent column vector concatenation. Each row  $C_j$  is a new feature representation generated from  $n$  filters and pooling for the window vector at position  $j$ , as indicated by the green dotted line box in Figure 9. Flatten  $\mathbf{C}$  into a one-dimensional feature sequence by row, and the new successive window feature sequence representations are then fed into the LSTM.

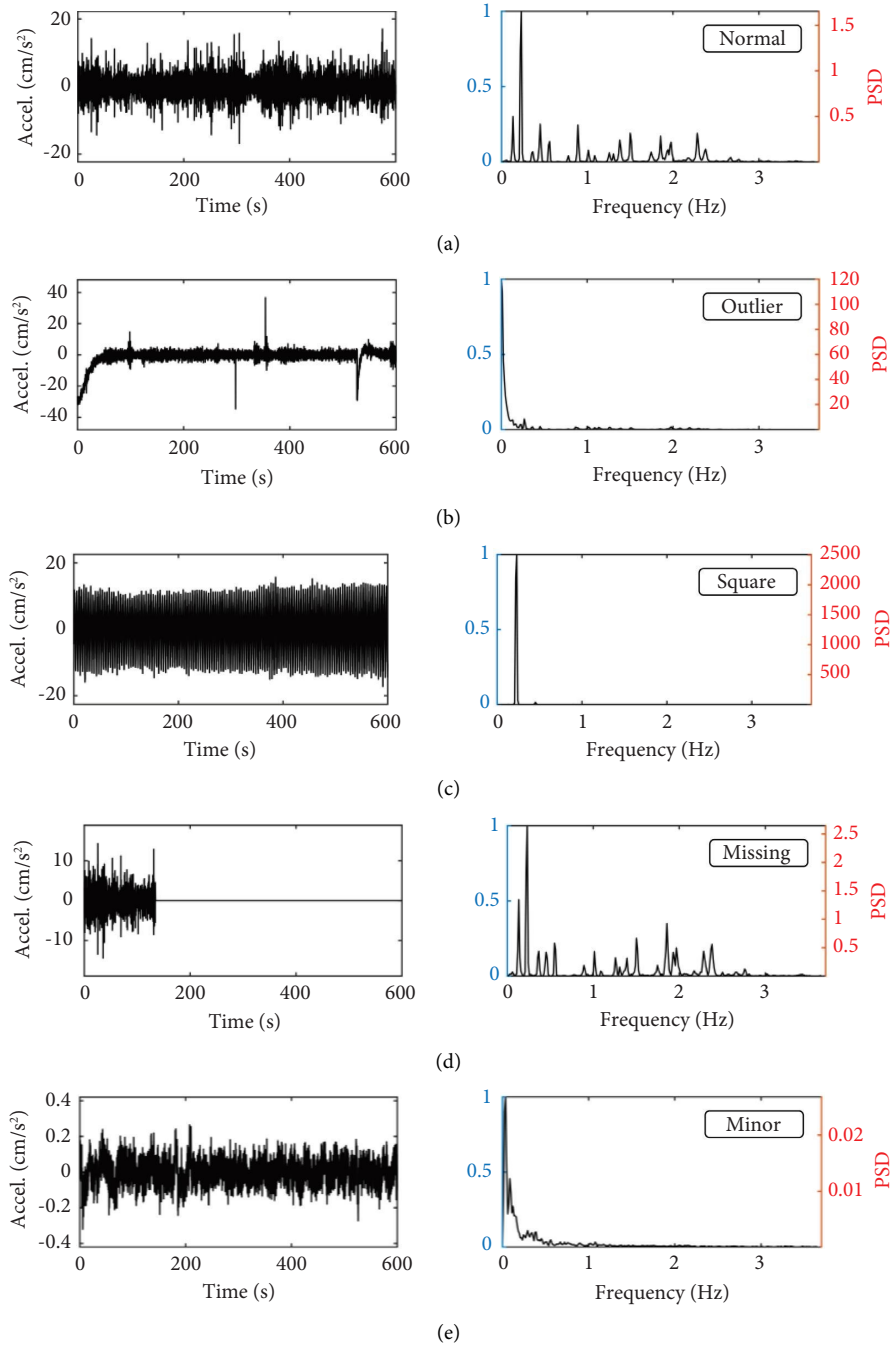


FIGURE 8: The power spectral densities of various data patterns: (a) normal; (b) outlier; (c) square; (d) missing; (e) minor.

**2.4.2. Long Short-Term Memory Networks.** LSTM [31] is a variant of recurrent neural network (RNN), as shown in Figure 10, and is used to solve the disadvantage that the standard RNN cannot learn long-term dependencies. The LSTM architecture has a range of repeated modules for each time step, similar to a standard RNN. At each time step, the output of the module is controlled by a set of gates in  $\mathbb{R}^d$  as

a function of the old hidden state  $h_{t-1}$  and the input at the current time step  $x_t$ : forget gate  $f_t$ , input gate  $i_t$ , and output gate  $o_t$ . These gates collectively decide how to update the current memory cell  $c_t$  and current hidden state  $h_t$ . We use  $d$  to denote the memory dimension in the LSTM, and all vectors in this architecture share the same dimension. The LSTM transition functions are defined as follows:



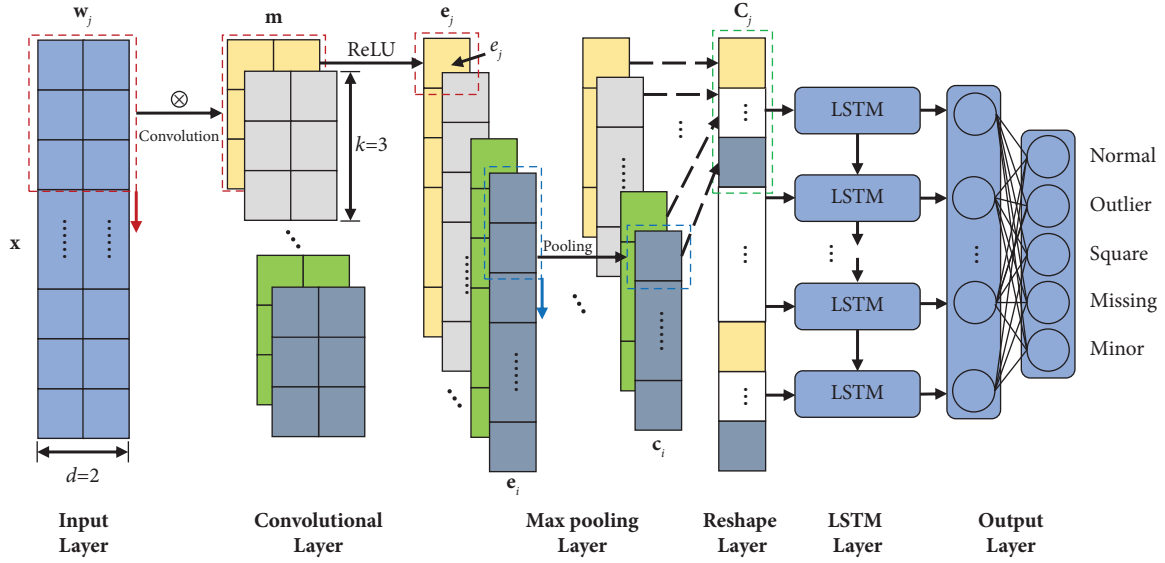


FIGURE 9: CNN-LSTM network structure.

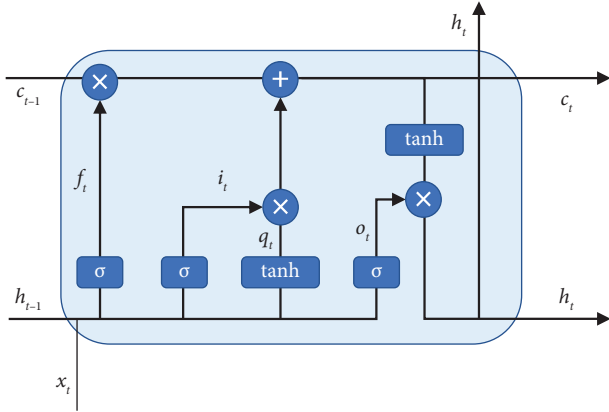


FIGURE 10: The architecture of the LSTM network.

$$\begin{aligned}
 i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\
 f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \\
 q_t &= \tanh(W_q \cdot [h_{t-1}, x_t] + b_q), \\
 o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\
 c_t &= f_t * c_{t-1} + i_t * q_t, \\
 h_t &= o_t * \tanh(c_t),
 \end{aligned} \tag{5}$$

where  $\sigma$  is the logistic sigmoid function with output in  $[0, 1]$ ,  $\tanh$  denotes the hyperbolic tangent function with output in  $[-1, 1]$ , and “\*” denotes element-wise multiplication. LSTM is explicitly designed for time-series data to learn long-term dependencies; therefore, we choose to use LSTM upon the convolution layer to learn such dependencies in the sequence of high-level features.

**2.4.3. Loss Function.** The loss function is applied to update the network parameters. It evaluates the classification accuracy of the network by measuring the error between the

discrete probability distributions of the real and predicted classes. The cross entropy is used as the objective function in this study and is defined as

$$H = -\frac{1}{N} \sum_{i=1}^N (p'_i \ln(p_i)), \tag{6}$$

$$p_i = \text{softmax}(q_i) = \frac{\exp(q_i)}{\sum_{l=1}^k \exp(q_l)},$$

where  $q_i$  denotes the  $i$ -th result of the output layer;  $k$  denotes the number of classification categories;  $p_i$  represents the prediction label of  $q_i$ , which is calculated using the softmax function;  $p'_i$  represents the  $i$ -th element in the real label and all labels are one-hot encoding; and  $N$  is the total number of samples.

**2.4.4. Performance Index.** In the statistical analysis of binary or multiple classifications, accuracy, precision, recall, and the  $F_1$  score are commonly used to measure the accuracy of network classification prediction:

$$\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}},$$

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

$$F_1 = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} * 100\%,$$

where TP denotes “true positive” (the actual value is true, and the classifier also predicted true); TN denotes “true negative” (the actual value is false, and the classifier also predicted false); FP denotes “false positive” (the actual value

is false, but the classifier predicted true); and FN denotes “false negative” (the actual value is true, but the classifier predicted false). The accuracy is generally used as an overall evaluation for all classes. Precision evaluates reliability based on the classification results. Recall can be regarded as a reliability evaluation based on the ground truth. Finally, the  $F_1$  score is the harmonic mean of precision and recall.

### 3. Example

**3.1. Bridge Overview.** The proposed method was applied to the structural monitoring data of a suspension bridge with a main span of 888 m in China. A permanent health monitoring system was installed on the bridge, including seven bidirectional acceleration sensors evenly installed on the upstream side to monitor horizontal (H1–H7) and vertical (V8–V14) vibrations. The other seven unidirectional sensors (V1–V7) were evenly arranged on the other side [32] to monitor only the vertical vibration of the bridge, as shown in Figure 11. The sampling frequency of the sensor is 50 Hz.

**3.2. Description of Datasets.** In this study, the acceleration data from 21 sensors for half a month (May 1–15, 2020) were utilized to build the anomaly detection training dataset. The original continuous data were divided at 10-minute intervals without overlapping windows, and the size of a single sample was  $1 \times 30000$ ; a total of 2160 sets of time-series measurements for each sensor were obtained. Samples were classified and marked based on their characteristics in the time and frequency domains. The examples in Figure 12 demonstrate that variations existed ubiquitously in all patterns. The middle of each category example in the figure is the original signal, the upper row is the residual signal used to characterize the time-domain feature, and the lower row is the PSD sequence used to characterize the frequency-domain feature. The actual dataset is severely imbalanced, which means that the number of normal samples far exceeds that of abnormal samples. Previous studies have shown that imbalanced datasets usually cause overfitting to the major classes and underfitting to the minor classes [16]. Therefore, the number of samples of outlier pattern and missing pattern can be expanded manually by randomly adding outliers to outlier pattern data and clearing part of normal pattern data, respectively, to avoid the impact of dataset imbalance on recognition accuracy, as shown in Figure 13. In addition, the number of normal samples was reduced to balance the proportions of various types in the dataset. Table 1 describes the number and features of normal data and the four types of abnormal data in the application dataset used in this study.

**3.3. Model Training and Experimental Results.** Python Science Suite, TensorFlow, and Keras were employed to build an improved CNN-LSTM model architecture with GPU acceleration. The processor and graphics card of the hardware platform were Intel Core i9-12900K and Nvidia GeForce RTX 3080Ti, respectively. According to the proposed method, anomaly detection was performed on the bridge monitoring dataset. First, the training and validation

sets were divided according to a ratio of 7 : 3 to constitute the final dataset. We trained the entire model by minimizing the cross-entropy error and used the Adam optimizer to automatically adjust the learning rate. After repeated tests, the model parameters were adjusted; the final model parameters are listed in Table 2. In addition, according to the size of the dataset used in this study, the number of training samples in each batch was determined to be 64, and the dropout rate was set to 0.5 to prevent overfitting. 12600 training samples were fed into the CNN-LSTM model for training. Define a stop trigger to determine the number of epochs. If the validation errors do not improve in the last 10 steps, the training can be stopped. The final model training was completed after 100 epochs. As the number of epochs increased, the overall training accuracy showed an upward trend; in contrast, the overall training loss showed a downward trend, as shown in Figure 14. After 100 epochs, the accuracy converged to approximately 0.996, and the loss value converged to approximately 0.014. The changes in the accuracy rate and loss value in the verification set are also shown in the figure.

Figures 15(a) and 15(b) show confusion matrices for inspecting the classification results, where diagonal elements are the numbers of correct classification results, and their corresponding recall rates are provided in brackets. In both the validation and training sets, the recall rate of each pattern reached more than 96%, and the total accuracies of the training and validation sets were 99.83% and 97.87%, respectively. Table 3 lists the results of various performance indexes of the model on the validation set. In terms of precision, the missing pattern had the highest accuracy at 99.8%, whereas that of the outlier and minor patterns was 96.83%. The square pattern had the highest recall rate and  $F_1$  score.

In Figure 16, the correctly classified samples in diagonal cells essentially display a stationary single feature according to the pattern, whereas the incorrectly classified samples in nondiagonal cells mostly display multiple features of the actual pattern and incorrect prediction patterns. For example, in  $M(4, 1)$  (the cell in the fourth row and the first column of the confusion matrix), the missing pattern samples are incorrectly predicted as normal pattern samples because they have good symmetry, few missing values, and frequency domain characteristics consistent with normal pattern data. In  $M(2, 1)$ ,  $M(2, 3)$ , and  $M(2, 5)$ , when the outlier pattern removes a few outliers, it shows the features of normal, square, and minor patterns, respectively. In  $M(5, 2)$ , there are very few outliers in the minor pattern, causing it to be incorrectly predicted as the outlier pattern. In addition, there were a few errors caused by manual classification, such as  $M(3, 1)$  and  $M(5, 1)$ .

## 4. Discussion

**4.1. Comparison with Other Methods.** To verify the improvement of the proposed anomaly vibration data detection model compared with existing methods, several existing models were additionally tested on the dataset of this study. When the model achieved the optimal result under multiple

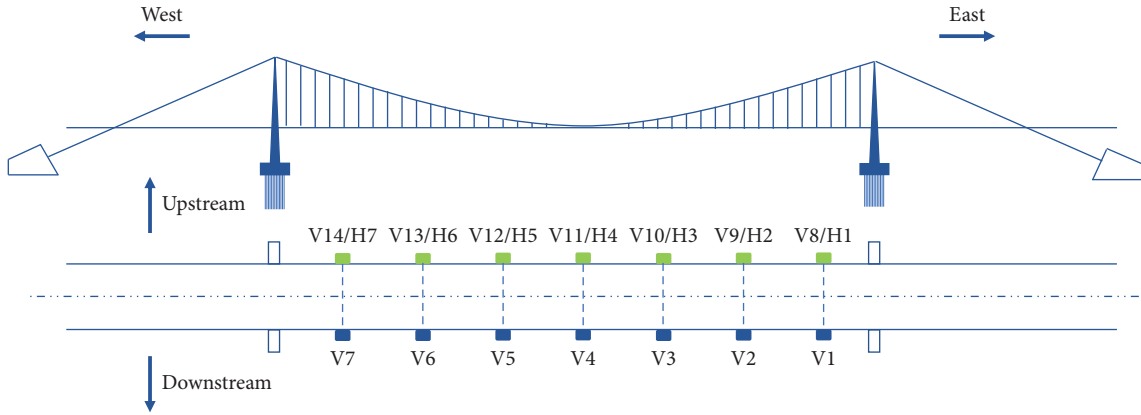


FIGURE 11: Arrangement of acceleration sensors on suspension bridge.

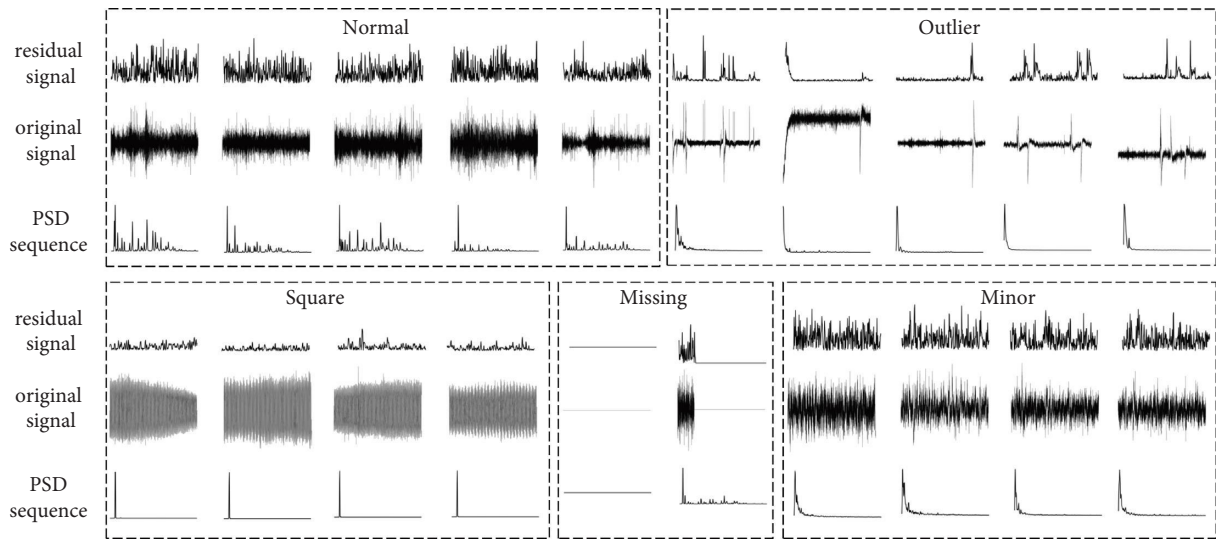


FIGURE 12: Examples of dataset grouped by class.

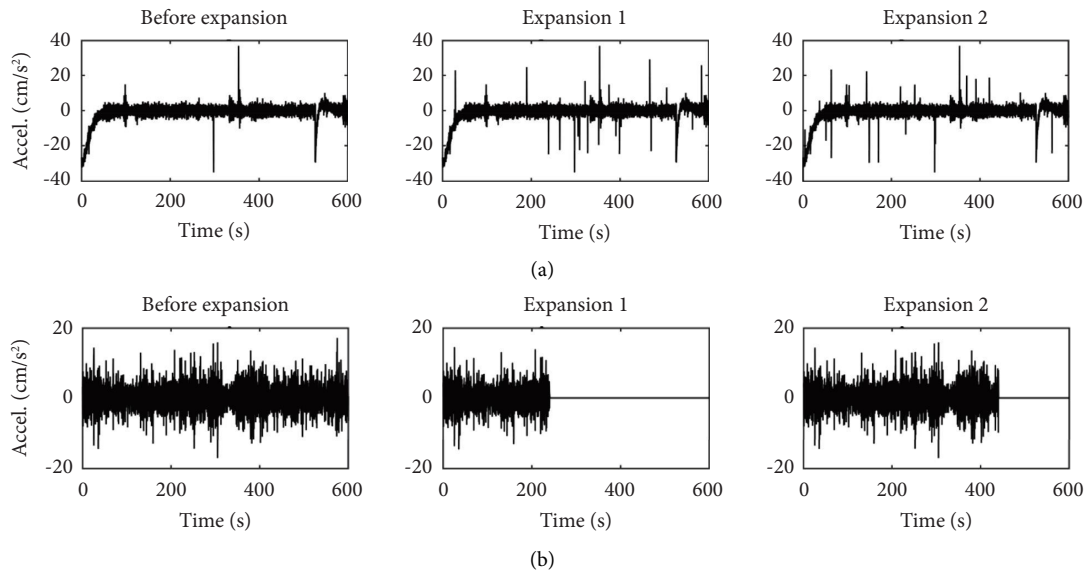


FIGURE 13: Comparison chart before and after data expansion: (a) data expansion for outlier data; (b) data expansion for missing data.

TABLE 1: Description of data patterns.

No.	Data patterns	Feature description	Quantity
1	Normal	The time response is a normal oscillation curve; the envelope has high symmetry; frequency response presents multiple peaks	3600 (20%)
2	Outlier	The time response has obvious outliers, trend, or drift; the envelope has no symmetry; the frequency response has no obvious peak	3600 (20%)
3	Square	The time response is like a square wave; the envelope is close to a horizontal line; the frequency response shows a single obvious peak	3600 (20%)
4	Missing	Most or all of the time responses are missing, most or all of the envelope values are zero, and the frequency response depends on the part that is not missing	3600 (20%)
5	Minor	The time response is a slight vibration; the envelope has no symmetry; the frequency response has no obvious peak	3600 (20%)

TABLE 2: Description of parameters of the LSTM-CNN model.

Network layer	Parameter setting	Activation function	Output size
Input layer	—	—	(None, 300, 2)
1D-convolutional layer	Size of convolution kernel: $3 \times 2$ Number of channels: 16 Stride: 1	ReLU	(None, 298, 16)
Max pooling layer	Pooling window: $2 \times 1$ Stride: 2	—	(None, 149, 16)
Reshape layer	—	—	(None, 2384)
LSTM layer	Number of nerve cell: 100	Tanh	(None, 100)
Fully connected layer	—	—	(None, 100)
Output layer	—	Softmax	(None, 5)

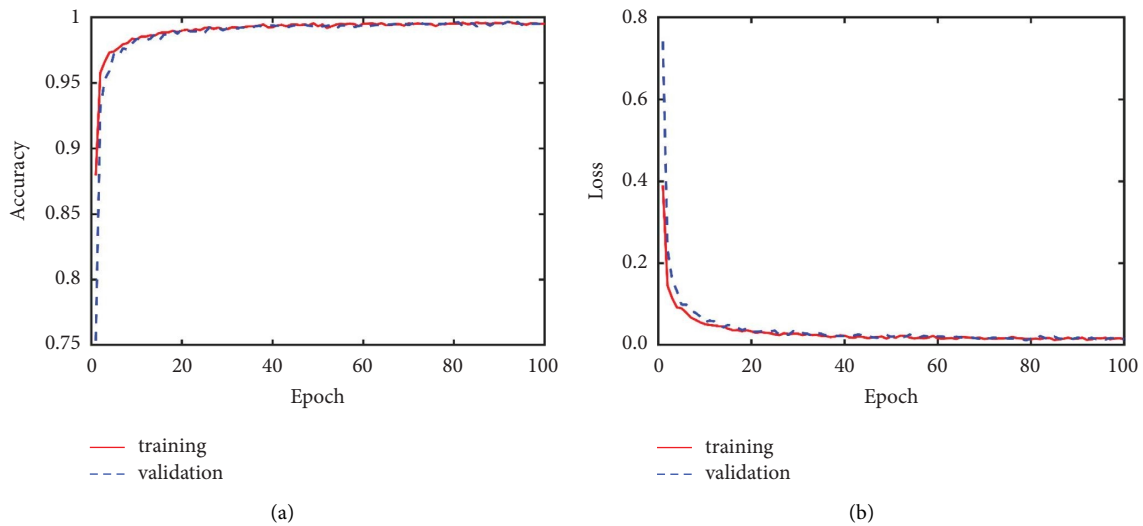


FIGURE 14: The change of accuracy rate and loss value: (a) accuracy rate; (b) loss value.

tests, the total accuracy, loss value, and total time consumption were selected as the measurement standards, as listed in Table 4.

The one-dimensional CNN model [21] consists only of a convolution layer and a fully connected layer, and the specific parameters are consistent with those in the proposed model. The total accuracy rate on the training set was only 88.78%; however, because of the simple network structure, the training time was the shortest. The CNN model [15, 16] consists of two layers of 2D convolution, two layers of 2D

pooling layer, and a fully connected layer. The input of the model was set as a  $128 \times 128$ -pixel grayscale image, which was constructed from the dataset of this study. The accuracy of the model was 89.43%, but the construction and importing of images consumed considerable time. For massive quantities of monitoring data, additional image data will also occupy a large amount of memory. As shown in Table 4, processing 18,000 pictures in the dataset took 583.3 seconds, whereas extracting the residual signal and PSD sequence of data took only 238.8 seconds.

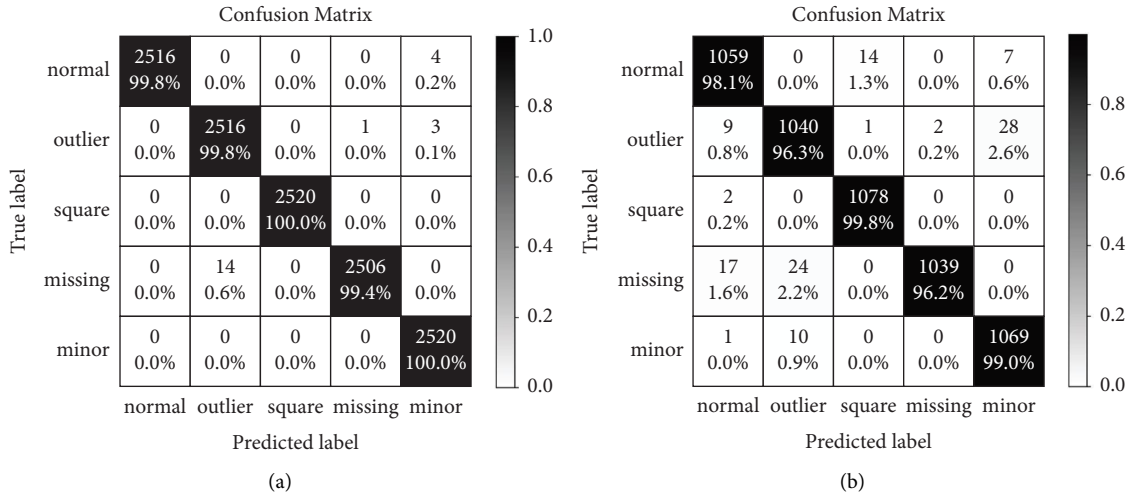


FIGURE 15: Confusion matrix: (a) training set; (b) validation set.

TABLE 3: Performance indexes of the LSTM-CNN model on the validation set.

Class	Anomalous patterns	Precision (%)	Recall (%)	$F_1$ score (%)
1	Normal	97.33	98.06	97.69
2	Outlier	96.83	96.30	96.56
3	Square	98.63	99.81	99.22
4	Missing	99.80	96.20	97.97
5	Minor	96.83	99.00	97.90
Accuracy		97.87%		

Moreover, the output of the fully connected layer of the two models was visualized through t-SNE [33]. As shown in Figure 17(a), the features of the same type of samples in the proposed method were relatively clustered, and the features of different types of samples were clearly distinguished. However, the features of the minor and normal patterns extracted by CNN, as shown in Figure 17(b), were severely confused owing to the visualized images, which did not contain the frequency-domain information and amplitude information of the data. As shown in Figure 18, it is difficult to classify these two classes solely by using images in the time domain. The LSTM model [19] and bidirectional LSTM (Bi-LSTM) model [20] were also used to realize time-series classification by connecting the fully connected layer, with accuracy rates of 89.02% and 94.46%, respectively. The training time of the Bi-LSTM model was more than twice that of the LSTM model. The CNN-LSTM model had the highest accuracy and lowest loss value. The front convolution and pooling layers shorten the size of the feature sequence, resulting in a shorter training time compared with that of the LSTM model. Therefore, compared with existing methods, the proposed method has higher efficiency and accuracy.

**4.2. Data Anomaly Detection for Long-Term Monitoring.** To further validate the performance of the proposed method on long-term monitoring data, the acceleration data of the suspension bridge collected from nontraining sets were continuously input into the program, and the trained model

was used to complete the detection of abnormal data. Figure 19(a) shows the distribution results of the data patterns detected by the model collected using the SHM system of the suspension bridge from May 16 to June 27, 2020. Evidently, these data patterns have certain spatial and temporal distribution rules in space and time. Normal data are the most widely distributed data and constitute the main data patterns. The outlier pattern is distributed on individual sensors and continuously distributed in a certain period of time, as shown in the blue part of H1 channel area 1 in the figure. The square pattern is displayed on all sensors or all vertical vibration sensors in a certain period of time, as shown in the yellow part of areas 2 and 3 in the figure. Finally, the missing pattern is similar to the square type. It appears on all sensors in a certain period, as shown in the black part of area 4 in the figure. The entire SHM system was temporarily down during this period, so the sensors did not record any data. The H3 channel of area 5 mainly includes the minor pattern. The slight vibration indicates that the sensor has been seriously damaged, and the data collected by it are clearly abnormal. In addition, these anomalies also occur sporadically at other times or on other sensors.

To verify the reliability of the proposed method, all data samples were manually detected and marked for comparison with the proposed detection results. The results are shown in Figure 19(b). The distributions of these abnormal data in time and space were consistent with the model detection results. Figure 20 shows the counting results of the model test results and actual data types. The recall rate for all patterns was above 90%. For the normal pattern,

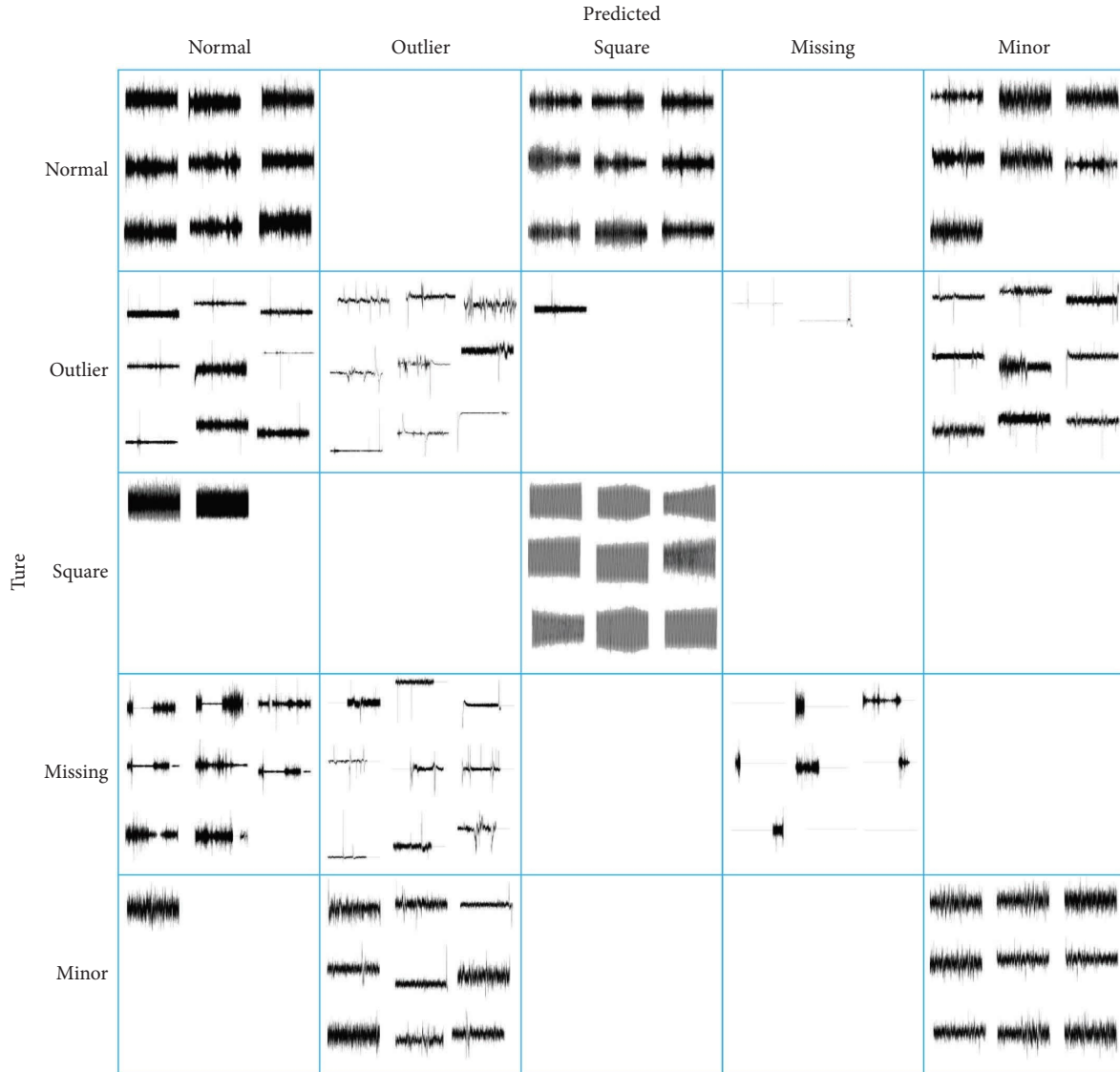


FIGURE 16: Examples in detection results arranged by confusion matrix.

TABLE 4: The calculation results of the different models.

Model	Accuracy (%)	Loss	Time (s)		
			Preprocessing	Training	Total
1D-CNN	88.78	0.5242	238.8	21.4	260.2
CNN	89.43	0.3740	583.3	54.9	638.2
LSTM	89.02	0.4158	238.8	86.6	325.4
Bi-LSTM	94.46	0.1256	238.8	200.0	438.8
CNN-LSTM	99.83	0.0140	238.8	37.8	276.6

0.7% of the normal samples were incorrectly classified as belonging to the minor pattern. 0.5% were incorrectly classified as the outlier pattern, 5.3% of outlier samples were incorrectly classified as minor patterns, 1.3% were incorrectly classified as missing patterns, and square patterns were mainly incorrectly classified as normal patterns. Of the missing patterns, 7.7% were incorrectly classified as outlier patterns, and 8.3% of the minor patterns were

incorrectly classified as outlier patterns. Therefore, the recall rates of these two patterns were lower than those of other patterns. The total accuracy rate of the data was 97.5%, and the average detection time of a single sample with a sampling time of 10 minutes was only 0.0365 seconds, which shows that this method can also efficiently and accurately detect various anomalies in long-term continuous monitoring data.



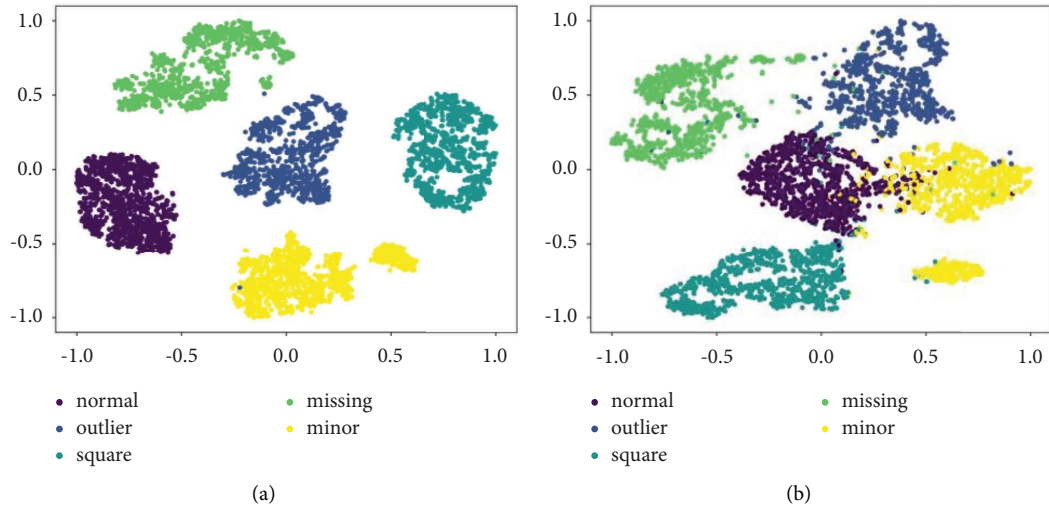


FIGURE 17: The t-SNE visualization of the output results of the fully connected layer on the validation dataset: (a) CNN-LSTM; (b) CNN.

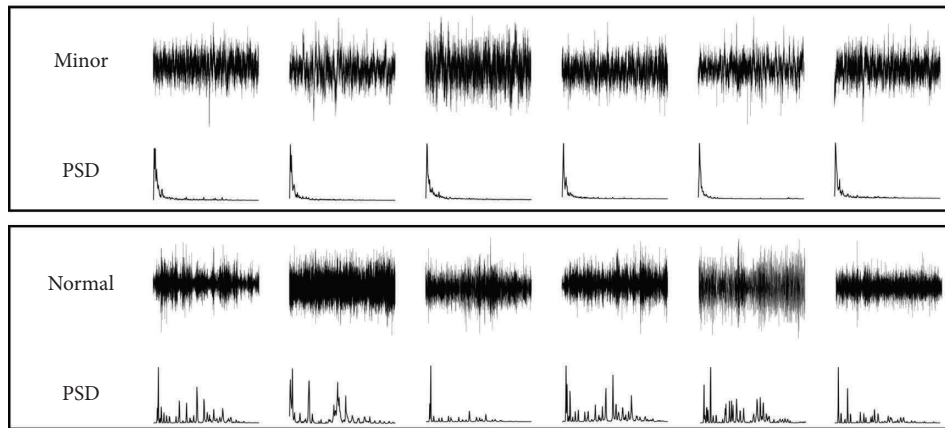
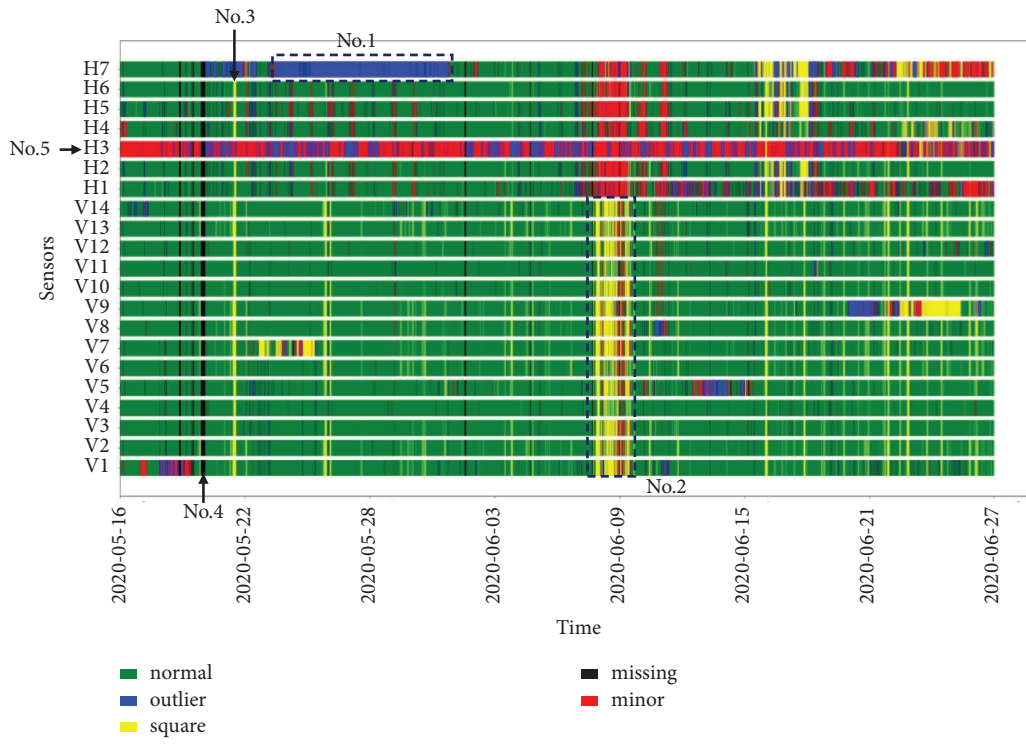


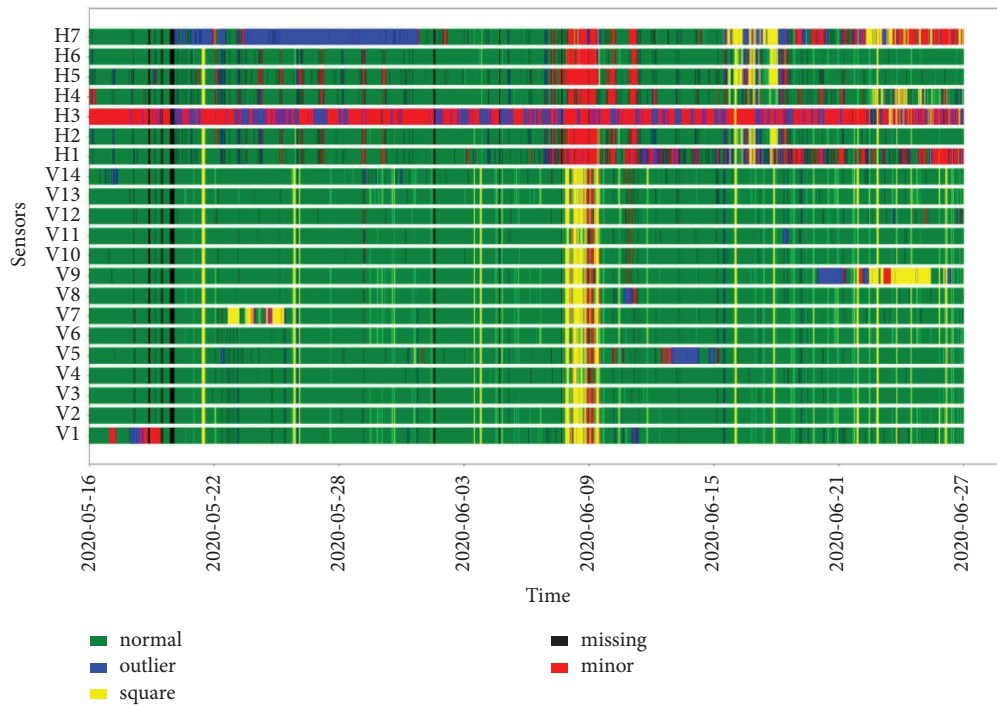
FIGURE 18: Visual comparison between minor pattern and normal pattern.

4.3. *Validation of Generalization Ability of the Proposed Method.* To validate the generalization ability of the proposed method, we employed vibration data from another SHM system of an actual long-span cable-stayed bridge to verify whether the trained model is effective for other structures. The system consisted of 19 channels. The sampling time was January 1, 2021, continuously covering 24 hours, with a sampling frequency of 50 Hz. The accelerometer positions of the bridge are shown in Figure 21.

The detection results obtained using the vibration data anomaly detection method proposed in this paper for this segment of data are shown in Figure 22. As shown in the figure, this method successfully detected a variety of abnormal data, and the detection results were consistent with the manual detection results, which shows that the proposed method has a strong generalization ability. It not only is effective on a single specific structure but also can accurately detect anomalies in the vibration data of other structures.



(a)



(b)

FIGURE 19: Distribution of data patterns from May 16 to June 27, 2020: (a) detection results; (b) actual data anomaly distribution.

Confusion Matrix

True label	normal	98105 98.5%	472 0.5%	262 0.3%	62 0.0%	697 0.7%
	outlier	30 0.4%	6861 93.0%	0 0.0%	99 1.3%	390 5.3%
	square	225 2.4%	0 0.0%	8973 97.5%	0 0.0%	3 0.0%
	missing	23 1.2%	143 7.7%	0 0.0%	1695 91.1%	0 0.0%
	minor	66 0.7%	743 8.3%	2 0.0%	0 0.0%	8157 91.0%
		normal	outlier	square	missing	minor
		Predicted label				

FIGURE 20: Confusion matrix between actual data and detection results.

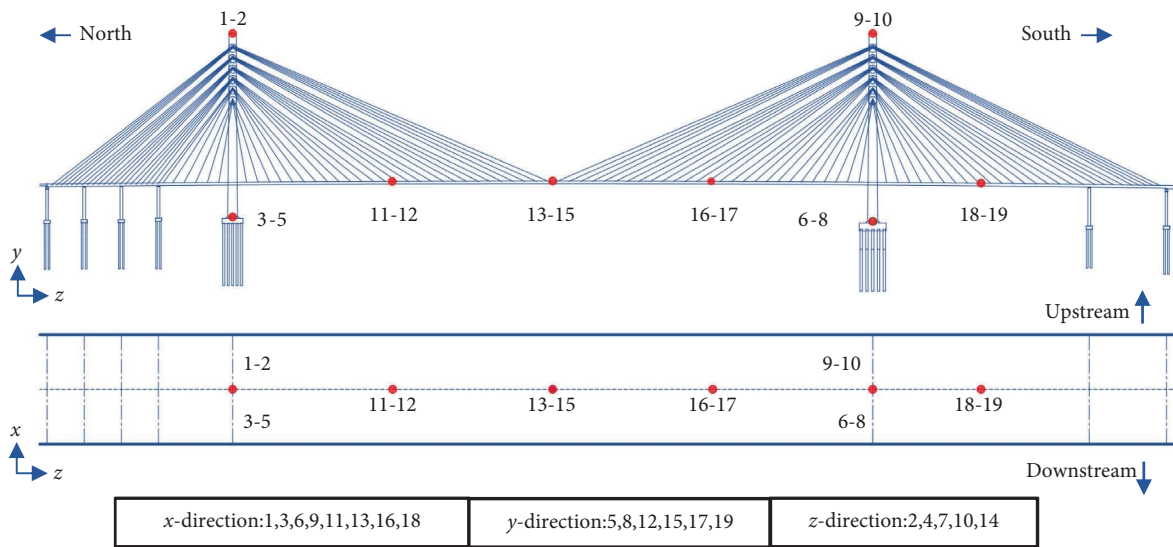


FIGURE 21: Sensor arrangement of cable-stayed bridge.

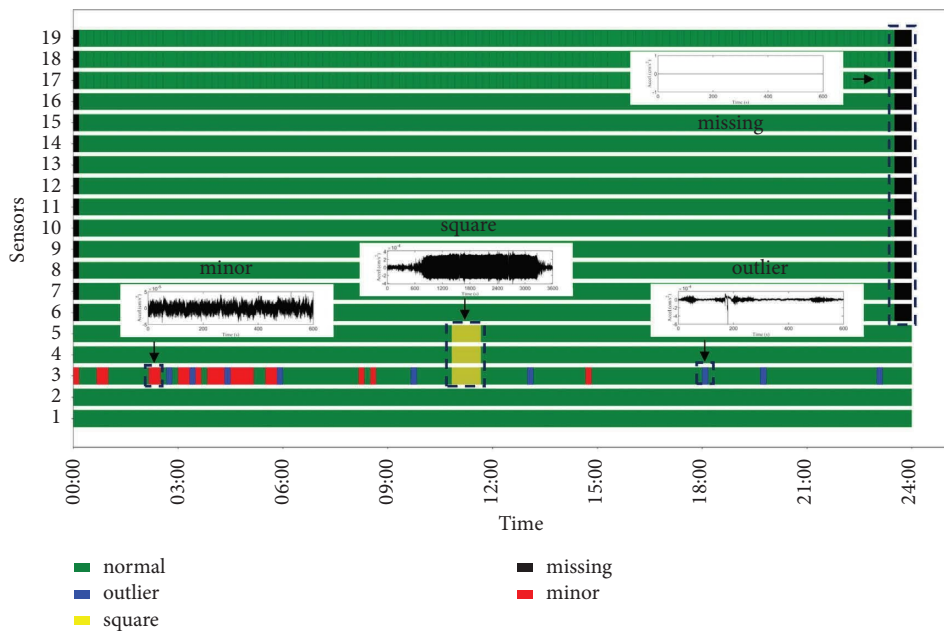


FIGURE 22: Detection results of abnormal vibration data of cable-stayed bridge.

## 5. Conclusions

In this study, the problem of abnormal vibration monitoring data detection was modeled as a standard time-series classification problem. The original vibration sequence was processed using feature engineering, and a feature sequence with a smaller size was obtained. Feature engineering includes the residual signal extracted from the upper and lower peak envelope of vibration data, which is used to characterize the symmetry of data in the time domain, and the power spectral density sequence of the data is used to characterize the interpretability of the data in the frequency domain. This study established a CNN-LSTM model for feature sequence classification. The CNN in the front of the model was used to extract the hidden features of the input sequence and construct a new time series that combines the features of the original signal in the time and frequency domains. The new feature sequences are then classified using LSTM and the fully connected layer. The effectiveness of this method was verified using the vibration monitoring data of a long-span suspension bridge. The results showed that the total accuracy of the training set and the validation set was 99.83% and 97.87%, respectively. The average accuracy on the test set composed of 42 days of continuous monitoring data exceeded 94%, and the average detection time of a single sample with a sampling time of 10 minutes was only 0.0365 seconds. The novelty of this framework is that it accurately imitates the manual preprocessing workflow of vibration monitoring data, including symmetry checks in the time domain and interpretability judgments in the frequency domain. In addition, dimensionless processing and downsampling operations are performed on feature sequences extracted by feature engineering, which increases the universality and computational efficiency of the method. Therefore, the proposed anomaly detection method can be extended to other structural types of vibration monitoring data and has significant advantages in detecting anomalies in long-term monitoring data.

## Data Availability

All datasets in this study were obtained by our research group during experiments, rather than publicly available datasets. The datasets used to support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This study was financially supported by the National Science Foundation of China (no. 51978217).

## References

- [1] S. Li, H. Li, Y. Liu, C. Lan, W. Zhou, and J. Ou, "SMC structural health monitoring benchmark problem using monitored data from an actual cable-stayed bridge," *Structural Control and Health Monitoring*, vol. 21, no. 2, pp. 156–172, 2014.
- [2] X. Zhang, W. Zhou, Y. Huang, and H. Li, "Automatic identification of structural modal parameters based on density peaks clustering algorithm," *Structural Control and Health Monitoring*, vol. 29, no. 12, p. e3138, 2022.
- [3] F. Zhang, Y. Ni, and H. Lam, "Bayesian structural model updating using ambient vibration data collected by multiple setups," *Structural Control and Health Monitoring*, vol. 24, no. 12, Article ID e2023, 2017.
- [4] A. Esfandiari, "An innovative sensitivity-based method for structural model updating using incomplete modal data," *Structural Control and Health Monitoring*, vol. 24, no. 4, Article ID e1905, 2017.
- [5] J. Chen and C. Loh, "Two-stage damage detection algorithms of structure using modal parameters identified from recursive subspace identification," *Earthquake Engineering & Structural Dynamics*, vol. 47, no. 3, pp. 573–593, 2018.
- [6] W. Zhou, S. Li, and H. Li, "Damage detection for SMC benchmark problem: a subspace-based approach," *International Journal of Structural Stability and Dynamics*, vol. 16, no. 04, Article ID 1640025, 2016.
- [7] E. Reynders and G. De Roeck, "A local flexibility method for vibration-based damage localization and quantification," *Journal of Sound and Vibration*, vol. 329, no. 12, pp. 2367–2383, 2010.
- [8] C. Rainieri, D. Gargaro, G. Fabbrocino et al., "Shaking table tests for the experimental verification of the effectiveness of an automated modal parameter monitoring system for existing bridges in seismic areas," *Structural Control and Health Monitoring*, vol. 25, no. 7, p. e2165, 2018.
- [9] G. Song, H. Li, B. Gajic, W. Zhou, P. Chen, and H. Gu, "Wind turbine blade health monitoring with piezoceramic-based wireless sensor network," *International Journal of Social Network Mining*, vol. 4, no. 3, pp. 150–166, 2013.
- [10] T. Yi, H. Huang, and H. Li, "Development of sensor validation methodologies for structural health monitoring: a comprehensive review," *Measurement*, vol. 109, pp. 200–214, 2017.
- [11] K. Thiyagarajan, S. Kodagoda, and L. Nguyen, "Predictive analytics for detecting sensor failure using autoregressive integrated moving average model," in *Proceedings of the IEEE Conference on Industrial Electronics and Applications*, pp. 1926–1931, Siem Reap, Cambodia, June 2017.
- [12] L. Li, G. Liu, L. Zhang, and Q. Li, "Sensor fault detection with generalized likelihood ratio and correlation coefficient for bridge SHM," *Journal of Sound and Vibration*, vol. 442, pp. 445–458, 2019.
- [13] H. Wan and Y. Ni, "Bayesian modeling approach for forecast of structural stress response using structural health monitoring data," *Journal of Structural Engineering*, vol. 144, no. 9, Article ID 04018130, 2018.
- [14] J. M. Estevez-Tapiador, P. Garcia-Teodoro, and J. E. Diaz-Verdejo, "Anomaly detection methods in wired networks: a survey and taxonomy," *Computer Communications*, vol. 27, no. 16, pp. 1569–1584, 2004.
- [15] Y. Bao, Z. Tang, H. Li, and Y. Zhang, "Computer vision and deep learning-based data anomaly detection method for structural health monitoring," *Structural Health Monitoring*, vol. 18, no. 2, pp. 401–421, 2019.

- [16] Z. Tang, Z. Chen, Y. Bao, and H. Li, "Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring," *Structural Control and Health Monitoring*, vol. 26, no. 1, Article ID e2296, 2019.
- [17] J. Mao, H. Wang, and B. Spencer, "Toward data anomaly detection for automated structural health monitoring: exploiting generative adversarial nets and autoencoders," *Structural Health Monitoring*, vol. 20, no. 4, pp. 1609–1626, 2020.
- [18] J. Zhang, J. Zhang, and Z. Wu, "Long-short term memory network-based monitoring data anomaly detection of a long-span suspension bridge," *Sensors*, vol. 22, no. 16, p. 6045, 2022.
- [19] K. Vos, Z. Peng, C. Jenkins, M. R. Shahriar, P. Borghesani, and W. Wang, "Vibration-based anomaly detection using LSTM/SVM approaches," *Mechanical Systems and Signal Processing*, vol. 169, Article ID 108752, 2022.
- [20] K. Yang, Y. Ding, H. Jiang, H. Zhao, and G. Luo, "A two-stage data cleansing method for bridge global positioning system monitoring data based on bi-direction long and short-term memory anomaly identification and conditional generative adversarial networks data repair," *Structural Control and Health Monitoring*, vol. 29, no. 9, p. e2993, 2022.
- [21] Y. Zhang and Y. Lei, "Data anomaly detection of bridge structures using convolutional neural network based on structural vibration signals," *Symmetry*, vol. 13, no. 7, p. 1186, 2021.
- [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] M. Arul and A. Kareem, "Data anomaly detection for structural health monitoring of bridges using shapelet transform," *Journal of Engineering*, vol. 8, p. p2003, 2022.
- [24] S. Sony, S. Gamage, A. Sadhu, and J. Samarabandu, "Vibration-based multiclass damage detection and localization using long short-term memory networks," *Structures*, vol. 35, pp. 436–451, 2022.
- [25] S. Sharma and S. Sen, "Real-time structural damage assessment using LSTM networks: regression and classification approaches," *Neural Computing & Applications*, vol. 35, no. 1, pp. 557–572, 2023.
- [26] Z. Ding, R. Hou, and Y. Xia, "Jaya-based long short-term memory neural network for structural damage identification with consideration of measurement uncertainties," *International Journal of Structural Stability and Dynamics*, vol. 22, no. 14, 2022.
- [27] Z. Ding, Y. Yu, and Y. Xia, "Nonlinear hysteretic parameter identification using an attention-based long short-term memory network and principal component analysis," *Nonlinear Dynamics*, vol. 111, no. 5, pp. 4559–4576, 2023.
- [28] Y. Zhang, X. Wang, Z. Ding, Y. Du, and Y. Xia, "Anomaly detection of sensor faults and extreme events based on support vector data description," *Structural Control and Health Monitoring*, vol. 29, no. 10, Article ID e3047, 2022.
- [29] C. Zhou, X. Cai, Y. Fu et al., "Tetraploid complementation proves pluripotency of induced pluripotent stem cells derived from adipose tissue," *Cell Proliferation*, vol. 48, no. 1, pp. 39–46, 2015.
- [30] V. Nair and G. Hinton, "Rectified linear units improve restricted Boltzmann machines vinod nair," in *Proceedings of the International Conference on International Conference on Machine Learning*, pp. 807–814, Haifa, Israel, June 2010.
- [31] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [32] D. Dan and H. Li, "Monitoring, intelligent perception, and early warning of vortex-induced vibration of suspension bridge," *Structural Control and Health Monitoring*, vol. 29, no. 5, p. e2928, 2022.
- [33] J. Zheng, Z. Jiang, and H. Pan, "Sigmoid-based refined composite multiscale fuzzy entropy and t-SNE based fault diagnosis approach for rolling bearing," *Measurement*, vol. 129, pp. 332–342, 2018.