Classification of External Vibration Sources through Data-Driven Models Using Hybrid CNNs and LSTMs

Liang, Ruihua; Liu, Weifeng; Kaewunruen, Sakdirat; Zhang, Hougui ; Wu, Zongzhen

Link to publication on Research at Birmingham portal

WILEY | Hindawi

Research Article

Classification of External Vibration Sources through Data-Driven Models Using Hybrid CNNs and LSTMs

Ruihua Liang [ID],[1,2] Weifeng Liu [ID],[1] Sakdirat Kaewunruen [ID],[2] Hougui Zhang [ID],[3] and Zongzhen Wu[4]

[1]*Key Laboratory of Urban Underground Engineering of Ministry of Education, Beijing Jiaotong University, Beijing 100044, China*
[2]*Department of Civil Engineering, School of Engineering, University of Birmingham, Birmingham B152TT, UK*
[3]*Institute of Urban Safety and Environmental Science, Beijing Academy of Science and Technology, Beijing 100054, China*
[4]*China Academy of Railway Sciences Corporation Limited, Beijing 100081, China*

Correspondence should be addressed to Weifeng Liu; wfliu@bjtu.edu.cn

Excessive external vibrations could affect the normal functioning and integrity of sensitive buildings such as laboratories and heritage buildings. Usually, these buildings are exposed to multiple external vibration sources simultaneously, so the monitoring and respective evaluation of the vibration from various sources is necessary for the design of targeted vibration mitigation measures. To classify the sources of vibration accurately and efficiently, the advanced hybrid models of the convolutional neural network (CNN) and long short-term memory (LSTM) network were built in this study, and the models are driven by the extensive data of external vibration recorded in Beijing, and the parametric studies reveal that the proposed optimal model can achieve an accuracy of over 97% for the identification of external vibration sources. Finally, a real-world case study is presented, in which external vibration monitoring was carried out in a laboratory and the proposed CNN+LSTM model was used to identify the sources of vibration in the monitoring so that the impact of vibration from each source on the laboratory was analyzed statistically in detail. The results demonstrate the necessity of this study and its feasibility for engineering applications.

1. Introduction

External vibration is a widely concerning issue in urban infrastructure construction. Excessive external vibration could affect the normal functioning of sensitive instruments such as electron microscopes in laboratories and could affect the integrity of historical artifacts in heritage buildings. In addition, excessive vibration and the radiated noise it generates could also affect the daily life of residents in near-field residential buildings. Therefore, it is necessary to measure and evaluate the external vibration for these sensitive buildings, which could facilitate the understanding of the magnitude of the vibration impact on the building, and also provide evidence for vibration mitigation design when necessary.

Activities that usually cause external vibration include the normal functions of rail transit lines, including overground and underground railway lines [1–3] and the operation of vehicles on roads, especially heavy trucks, wagons, and buses [4]. Moreover, it also includes construction activities, especially impacts pile driving [5]. The vibrations generated by these activities could transmit through the soil to the surrounding area and generate vibration effects on the buildings in these areas.

The general vibration mitigation methods include (i) vibration mitigation at the source, such as using highly damped track structures to mitigate train-induced vibration [6, 7], (ii) vibration mitigation at the propagation path, such as implementing vibration isolation piles in the ground between the external excitation and the building [8, 9], and (iii) passive vibration isolation in sensitive buildings, such as installing vibration isolation tables under sensitive instruments and historical artefacts [10]. The vibration isolation measures for sensitive buildings require targeted

design according to the specific vibration source type and the magnitude of vibration it generates.

In practical engineering, there are usually multiple vibration sources working simultaneously around sensitive buildings. However, it is often unable to observe and record the exact time of operation of each vibration source in vibration monitoring, especially for underground vibration sources [11]. These situations make it difficult to statistically obtain the magnitude of the impact of each external excitation, and thus to take accurate and targeted vibration mitigation designs. Therefore, identifying the source of the measured vibration becomes a necessary part of the external vibration monitoring process.

Due to the distinct mechanisms, the vibration signals generated by various vibration sources show differences in characteristics in both the time and frequency domains. Experienced experts can identify the source of the vibration signals based on these characteristics and then carry out subsequent vibration evaluations. Nevertheless, a number of researches in recent years have shown that the magnitudes of external vibrations measured at different times often exhibit large differences [12, 13], which means that longer vibration monitoring than before is needed to accurately understand the magnitude distribution of external vibration in sensitive buildings. However, it is very inefficient and unfeasible to still identify the vibration sources by manual methods during the long-time vibration monitoring process.

In recent years, deep learning is becoming an efficient and well-known technique and has been widely used in the research and engineering of mechanical signal processing and identification [14–16], which can significantly reduce the need for manual labor and increase the accuracy of signal identification tasks. Thus, the use of deep learning methods to identify the source of external vibrations is a promising concept that can greatly improve the efficiency of external vibrations monitoring and evaluation. Compared with most existing studies on the identification of mechanical vibration, the difficulty in identifying external vibration lies in the high uncertainty of the vibration signal, including the uncertainty of the distance between the source and the measurement point and the uncertainty of the source parameters [17, 18]. Therefore, the effectiveness of deep learning-based external vibration source identification needs to be investigated in detail before applying the technique to practical projects.

In signal processing using deep learning methods, the most well-known models include a convolutional neural network (CNN) [19], long short-term memory (LSTM) network [20], and deep residual network [21]. Specifically, CNN can focus on the local features in the signal, while LSTM can focus on the sequence features in the signal, and both of them have achieved great success in signal identification tasks, such as damage identification of mechanical components [22, 23] and structural health monitoring [24–26]. In recent years, many researchers have used hybrid models based on both models to conduct research in the field of image or time series signal processing, such as daily energy consumption prediction [27, 28], daily air and water quality prediction [29, 30], financial asset price volatility prediction [31, 32], and biological and structural health monitoring [33–35]. These studies have demonstrated that hybrid models can exploit the strengths of each submodel simultaneously, achieving better results than any single model. Nevertheless, there are actually several forms of hybrid models, for example, combining CNNs and LSTMs in parallel, or combining CNNs and LSTMs in series. There are few existing studies that discuss the performance differences between the various forms of hybrid models, and this needs to be further investigated. Therefore, this paper aims to build various forms of hybrid models of CNN and LSTM to study their feasibility and effectiveness in the task of external vibration source identification.

Specifically, in this paper, hybrid models based on CNN and LSTM with various structures are developed, a large number of external vibration signals measured in Beijing are used as training and testing samples, and the most suitable model for external vibration source identification is investigated. Moreover, a case study is carried out to further discuss the performance and the feasibility of the proposed model in actual engineering, and this case study evaluates the vibration level in a laboratory as a preliminary step to propose isolation measures

The rest of the paper is organized as follows: Section 2 introduces the theoretical fundamentals of the data preprocessing methods and the deep learning models. Section 3 presents details of external vibration monitoring and signal preprocessing. Section 4 presents the implementation and performance of the proposed models. Section 5 presents a case study to demonstrate the feasibility of the proposed method in external vibration evaluation, and Section 6 concludes.

2. Related Theoretical Knowledge

2.1. Time-Frequency Analysis Techniques. In the field of deep learning-based signal identification, inputting time-domain signals directly into the model is the most common way of signal processing. However, the characteristics of the vibrations generated by different sources differ in both the time and frequency domains. Therefore, this paper uses time-frequency analysis techniques to preprocess the original monitored vibration signals into time-frequency spectra, to ensure that the deep learning model can take into account the characteristic differences of the vibration from each source in both time and frequency domains.

The most well-known time-frequency analysis techniques used in existing studies include the short-time Fourier transform (STFT), continuous wavelet transform (CWT), and Hilbert–Huang transform (HHT). Considering that the time-frequency spectrum obtained from the STFT exhibits different dimensions at different frequencies, which is not convenient for subsequent input to the deep learning model so that only CWT and HHT are considered for the analysis in this paper.

2.1.1. CWT. CWT is a renowned time-frequency analysis technique that provides promising performance in the time-frequency analysis of transient and nonstationary signals,

and it has been used in various fields of research for signal feature extraction [36]. The process of CWT analysis can be defined as follows:

$$\text{CWT}_\psi(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t)\psi\left(\frac{t-b}{a}\right)dt, \quad (1)$$

where a denotes the scaling coefficient, which is used to characterize the wavelet expansion; b denotes the shift coefficient, which is used to guide the location of the wavelet; $s(t)$ is the external vibration signal to be analyzed; and $\psi(t)$ is the wavelet basis function. The common basis functions used in the time-frequency analysis of vibration signals include Morlet (Morl), Mexican Hat (Mexh), and complex Gaussian (Cgau).

2.1.2. HHT. The HHT is another time-frequency analysis technique specifically developed to deal with nonlinear and nonstationary data [37]. Compared to other traditional methods, this technique can adaptively generate basis functions, so that the time-frequency spectrum obtained using this technique exhibits higher resolution, and it has also been widely used in engineering research in recent years. Briefly, HHT is a combination of empirical mode decomposition (EMD) and Hilbert transform. First, EMD allows any complex signal to be adaptively decomposed into intrinsic mode functions, which can be expressed as follows:

$$s(t) = \sum_n^{i=1} C_i + r_n, \quad (2)$$

where C_i is the ith decomposed intrinsic mode function, r_n is the residual signal. Then, the Hilbert transform is done for each intrinsic mode function, and the Hilbert spectrum $H(\omega, t)$ for the original signal can be expressed as follows:

$$H(\omega, t) = \text{Re} \sum_{i=1}^{n} a_i(t)e^{j\int \omega_i(t)dt}, \quad (3)$$

where Re is the operator of the real part, $a_i(t)$ and $\omega_i(t)$ are the amplitude and instantaneous frequency functions of each intrinsic mode function C_i. A more detailed description of this technology can be found in the existing studies [37, 38].

2.2. Hybrid Machine Learning Approach. This Section introduces the theoretical fundamentals of CNN, LSTM, and hybrid models of CNN and LSTM, which are deep learning algorithms that have been widely proven to be very effective in vibration analysis.

2.2.1. CNN. CNN is a well-known deep learning architecture that is widely used in research on the classification and identification of one-dimensional and two-dimensional signals. Due to techniques such as receptive fields, weight sharing, and pooling [39], CNN networks achieve lower complexity and better generalization performance than fully connected neural networks [40]. Well-known CNN models

such as ResNet [41], AlexNet [19], and GoogleNet [42] have all performed accurately on large-scale datasets in international competitions.

CNNs are usually composite networks with multiple stacked layers, the main type of layers in CNNs consists of convolutional layers, pooling layers, and fully connected layers. The illustration of the CNN model is shown in Figure 1. The signal fed into the model is first processed in the convolutional and pooling layers for feature extraction, and the extracted features are further integrated into the fully connected layer and finally mapped to the output layer. Specifically, the data processing of the signal in each type of layer can be represented as follows:

(1) Convolutional layer

The operation in the convolutional layer is given by

$$\mathbf{y}_n = f(\mathbf{ck}_n \otimes \mathbf{Y} + \mathbf{b}_n), n = 1, \ldots, N, \quad (4)$$

where $\mathbf{y}_n$ denotes the nth features extracted in the convolutional layer; $\mathbf{ck}_n$ is the nth convolution kernel; $\mathbf{b}_n$ is the nth bias; $\mathbf{Y}$ is the signal fed in the convolutional layer; N denotes the number of convolution kernels in the convolutional layer; $\otimes$ denotes the convolution operation; and $f(\cdot)$ denotes the nonlinear activation function, e.g., the rectified linear unit (ReLU) [43].

(2) Pooling layers

The pooling layer could reduce the size of the feature map extracted from the last convolutional layer, and pick out the most important features [19]. Common pooling methods include average pooling and max pooling, the process of max pooling could be described as follows:

$$\overline{\mathbf{y}}_n = \max\{\mathbf{y}_n\}, \quad (5)$$

where $\overline{\mathbf{y}}_n$ denotes the feature with reduced size; $\mathbf{y}_n$ denotes the extracted feature from the convolutional layer.

(3) Fully connected layer

The extracted features are then fed into the fully connected layer for classification, the fully connected layer is the main component of the fully connected network (FCNN). On this layer, all extended one-dimensional feature vectors are weighted, summed, and entered into the activation function, which could be described as follows:

$$y^i = f\left(w^i x^{i-1} + b^i\right), \quad (6)$$

where i is the index of the network layer, y is the output of a fully connected layer, x is the feature vector, w is the weight coefficient, and b is the bias. The softmax function is commonly used as the activation function $f(\cdot)$ for classification tasks, it can map the input to the probability of each category, which is defined as follows:
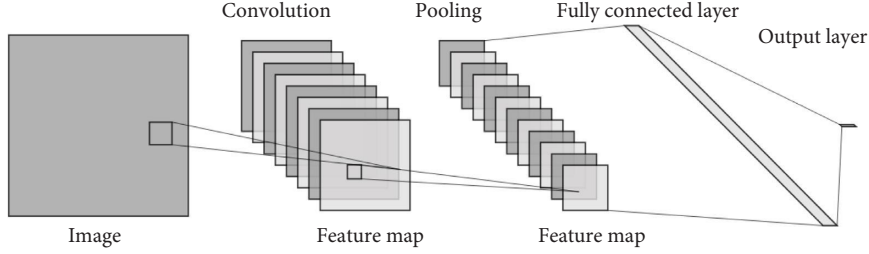
FIGURE 1: Illustration of the CNN model structure.

$$\varphi(r_c) = \frac{e^{r_c}}{\sum_{c=1}^{C} e^{r_c}}, c = 1, \ldots, C, \qquad (7)$$

where $\varphi = [\varphi(r_1), \ldots, \varphi(r_C)]$ denotes the probability of each classified result, r_c denotes the input of the activation function, and C denotes the number of signal categories.

Considering the nonperiodicity and nonstationarity of vibration signals, the batch normalization (BN) technique is usually used when processing such signals with CNN models. The BN layers are usually added following each convolutional layer, and it ensures that the features can maintain the same distribution during model training by fixing the mean and variance of the input signal for each convolutional layer, which can thus improve the model training efficiency [44].

In addition, to prevent overfitting, the dropout technique [45] is often used in fully connected layers, which can prevent over-reliance of the model on certain weights.

2.2.2. LSTM. LSTM [20] is a variant of recurrent neural network (RNN), which is a well-known neural network that can be used in the analysis of sequence data, such as analysis of language, vibration, and image. In contrast to traditional RNNs, LSTMs introduce memory units and gating mechanisms to deal with the gradient disappearance and gradient explosion problems during the training of long sequences. The gating mechanism in LSTM can be used to control the transfer state, aiming to remember the important information and forget the unimportant information. These functions are performed by the input, forget, and output gates in the network unit, and the unit of the LSTM is shown in Figure 2, the specific formulas of the LSTM are as follows:

$$
\begin{aligned}
f_t &= \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right), \\
i_t &= \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right), \\
\overline{c}_t &= \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_c\right), \\
c_t &= f_t \cdot c_{t-1} + i_t \cdot \overline{c}_t, \\
o_t &= \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right), \\
h_t &= o_t \cdot \tanh\left(c_t\right),
\end{aligned}
\qquad (8)
$$

where x_t is the input vector; f_t, i_t, and o_t are the activation vector of the forget gate, input gate, and output gate at time t, respectively; c_t and h_t are the memory cell activation vector

and hidden state activation vector at time t; W_f, W_i, and W_o are input kernels for the forget gate, input gate, and output gate, respectively; and b_f, b_i, b_c, and b_o are bias; σ is the logistic sigmoid function; tanh is the hyperbolic tangent activation function.

2.2.3. Integration of CNN and LSTM. The hybrid model of CNN and LSTM can be expressed in three forms, as shown in Figure 3. The first one can be called CNN–LSTM [26, 46], which indicates that the CNN is used to extract the local features of the signal first and the LSTM is used to further process the extracted features. The second one can be called LSTM–CNN [47, 48], which means that the LSTM is first used for the extraction of the overall features of the signal, and the CNN is subsequently used for further processing of the signal. The third type can be called CNN+LSTM [49, 50], in which both CNN and LSTM are used for feature extraction of the original input signal. In all three types, the obtained features are usually fed into a fully connected layer and classifier for signal classification and identification.

2.2.4. Hyperparameter Optimization. There are various hyperparameters in each submodule of the hybrid model of CNN and LSTM, such as the number of convolutional kernels and the number of LSTM units, and variations in these parameters during the training process could influence the accuracy of the model. To achieve the best identification results as far as possible, it is necessary to optimize the combination of hyperparameters in the model.

Common hyperparameter optimization methods can be divided into empirical and model-based methods [51]. Empirical hyperparameter optimization methods mainly refer to researchers relying on previous experience to adjust the hyperparameters of the model, which is accessible and has been applied by a large number of studies [27, 30, 31]. However, the empirical methods lack theoretical analysis and the optimal hyperparameter might be missed. Model-based methods usually include grid search and random search.

Specifically, grid search is an exact hyperparameter optimization method that tests all potential combinations of hyperparameters to obtain the optimal one [52, 53]. In grid search, the lower and upper values of each hyperparameter need to be artificially defined so that a parameter grid can be built up in the defined parameter areas, and the optimal hyperparameter combination is obtained by evaluating all
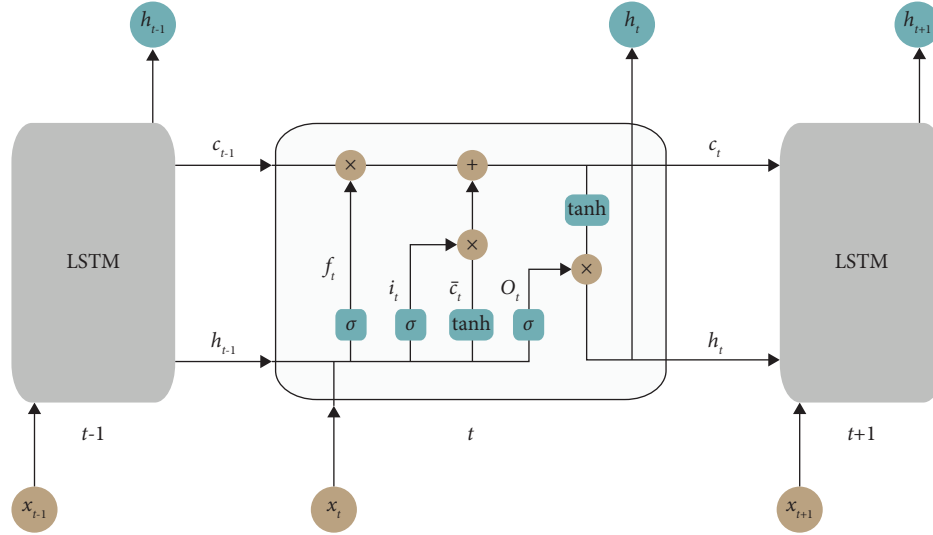
Figure 2: Illustration of the LSTM model.

the target function values for each node. The algorithm is simple to implement and effective in low-dimensional parameter spaces. However, it is computationally expensive and might not be suitable in cases where the model contains a large number of hyperparameters or where the model is computationally expensive.

Another universal method of hyperparameter optimization is random search, in which upper and lower limits for all hyperparameters also need to be artificially specified, but the method assumes that different hyperparameters are not equally important for the model performance, so it is more efficient to use randomly distributed points rather than uniformly distributed points to cover the parameter space [51, 54]. In practice, it has been found that random search is also easy to implement and performs better than grid search, because it can find the optimal combination of hyperparameters almost as well as grid search and in significantly less time [51]. Considering that the hybrid model of CNN and LSTM proposed in this study is computationally expensive, the random search is used for hyperparameter optimization in this paper.

2.3. Evaluation Indicators for Classification Performance. To find the best performing hyperparameter combinations for various models, the identification performance of various models with different hyperparameter combinations on the test set was evaluated using the indicator of accuracy, which was defined as

$$\text{Accuracy} = \frac{N_C}{N_T}, \tag{9}$$

where N_C is the number of samples correctly identified by the model in the test set, and N_T is the total number of samples in the test set.

In addition, to study the specific performance of the proposed method in the identification of various categories of vibration signals, the indicators of precision and recall are used in this paper, which are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{10}$$

where TP is the number of true positives, which is the positive vibration samples of a certain category that are correctly identified by the proposed model; FP is the number of false positives, which is the positive vibration samples of the certain category that are incorrectly identified by the proposed model; FN is the number of false negatives, which is the negative vibration samples of the certain category that are incorrectly predicted by trained model. Especially, precision indicates the percentage of samples identified correctly in the samples identified as positive vibration of a certain category by models, while recall indicates the percentage of samples identified correctly in the positive vibration samples of a certain category.

For a well-trained model, both performance indicators of precision and recall are expected to be high, but these two are contradictory to each other in some cases and it is difficult to improve both indicators at the same time. Therefore, this study also uses the F_1 score as the indicator of model performance, which considers both precision and recall and conveys a balance between them, and a well-trained model are expected to have a high F_1 score. The F_1 score is calculated as follows:

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{11}$$

3. External Excitations, Field Measurements, and Signals Preprocessing

This section presents the details of the external vibration signals used in the model training and testing in this study, including the description of the external vibration monitoring and signals preprocessing.
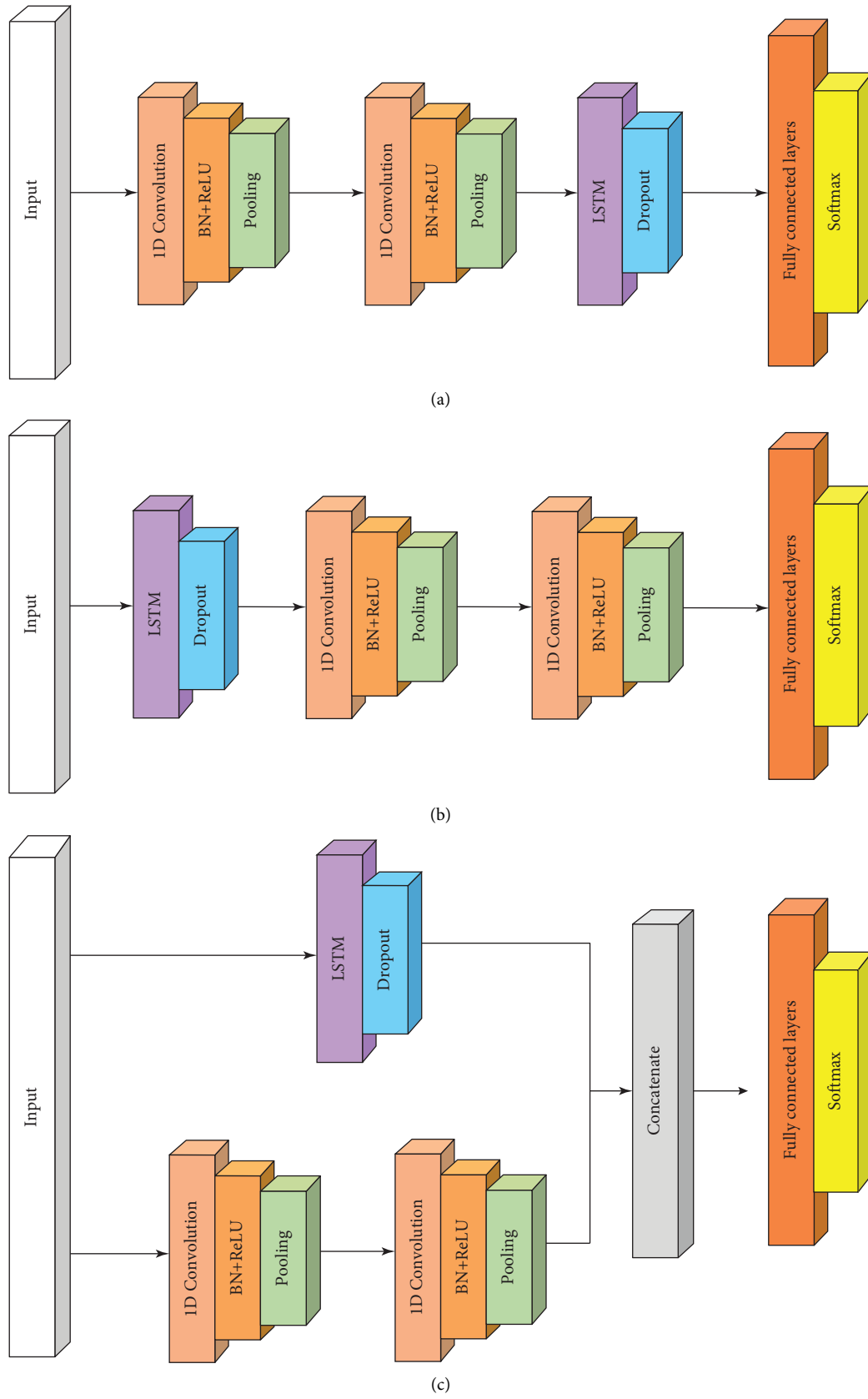
(a)

(b)

(c)

Figure 3: Hybrid structure of CNN and LSTM: (a) CNN − LSTM, (b) LSTM − CNN, and (c) CNN + LSTM.

3.1. External Excitations and Field Measurements. Figure 4 shows the process of external vibration monitoring in sensitive buildings. According to the type of sensitive building and the demand for vibration monitoring, multiple monitoring points can usually be set up inside and outside the building, and data acquisition instruments and vibration sensors are used to record the vibration signals. In actual projects, the sensors typically could record external vibrations from a variety of sources, including railway traffic, road traffic, and construction activities.

One of the most common sources of external vibration is vehicles on the road, the magnitude of which is closely related to the weight and speed of the vehicle. Trucks and buses with large loads usually generate excessive vibration, which is one of the external excitations of wide concern [55]. Since the type of vehicles driving on roads and the congestion of the traffic are random, this leads to high randomness of the ground vibrations generated by road traffic. In addition, with the extensive planning and construction of rail transit infrastructure around the world, the effects of railway-induced ground vibrations have been widely concerned. Although the physical parameters and driving conditions of trains passing through the same location are usually similar, recent studies have shown that the vibration induced by railway is also highly random due to the randomness of wheel wear [17]. Also, the weight of the train vehicle is significantly larger than that of the vehicle on the road, which results in a larger magnitude of ground vibration, making it one of the most concerning sources of vibration in sensitive buildings. Construction activity is also a source of external vibration that has been widely considered. There are many specific sources of vibration generated during construction activities, of which the impact pile driving is one of the external excitations of wide concern, which can generate impulsive excitations of large magnitude and the vibrations generated can propagate over long distances on the ground [56]. Therefore, the impact pile application is regarded as the major source of vibration for construction activities to be considered in this study.

Over the past decade, the author's research team has carried out extensive monitoring of external vibration at various sites in Beijing, and the locations of these monitoring sites are shown in Figure 5. These existing monitoring data provide sufficient samples for the study of deep learning-based vibration source identification in this paper, including samples for training the models and testing their effectiveness.

Specifically, the sensors used for this monitoring are the Lance accelerometer LC0130 with a measurement range of 0.12 g and measurement frequency range of 0.5–1000 Hz, the INV3060S networked distributed data acquisition devices were used to collect external vibration data, and computers were used to record and process the collected vibration data, these equipment are commonly used in external vibration monitoring, which is shown in Figure 6. In addition, multiple monitoring points were usually arranged at each monitoring site, which was usually placed on the ground at a distance of 0.5 m from the sensitive building, which satisfied the requirements of the relevant standards. Moreover, during the monitoring process, the technicians recorded the surrounding traffic and construction activity at each moment, this information assisted the experts in identifying the source of the vibration response in the subsequent processing.

3.2. Signal Preprocessing. In actual external vibration monitoring, only signal segments with high magnitude are usually of interest, which is potentially influential to sensitive buildings. Therefore, in this study, the original long-term monitored signals are intercepted and segmented, the length of the intercepted signal segment is chosen to be 30 seconds, which can usually contain most of the intact vibration signal caused by external excitation, and only those signal segments with the magnitude of the vibration root mean square (RMS) value higher than a certain threshold are selected for identification analysis, and the threshold needs to be taken according to the vibration limit value of the sensitive building to be evaluated. In this paper, this threshold has been chosen as 0.001 m/s^2, which is less than the limit value for urban ambient vibration according to Chinese standards [57], and the illustration of the segmentation of vibration signals in this paper is shown in Figure 7.

Another essential step in signal preprocessing is the labeling of the source of the vibrations in each signal segment. In this process, experts with experience in vibration evaluation identify the sources of vibration in each signal segment and use the identification results to label the signal segments. Specifically, the evidence for the identification by the expert mainly comes from two components. The first component is the recorded observation of the vibration source around the measurement point at each moment during the monitoring process, which is usually done by the technician who carried out the vibration monitoring. The other part is the characterization of the signal segment obtained in the analysis in the time and frequency domains. In this paper, vibration sources for signal segments were classified into four categories, including railway, road vehicles, construction activities, and other sources of vibration. The first three categories are the main sources of external vibration mentioned above, while other sources include environmental vibrations, mechanical vibrations such as those caused by ventilation equipment, and other vibrations that are not in the first three categories. The magnitude of these vibrations is usually relatively insignificant and is not of priority concern in external vibration monitoring.

It should be noted that the same signal segment may contain vibration signals from multiple sources, but only the source of the vibration with the largest magnitude was used as the label for the signal segment in the analysis in this paper. This allows for the identification of the most significant external sources of vibration and thus facilitates the vibration evaluation and design of the building for vibration mitigation. Specifically, according to the segmentation method above, the vibration magnitude at the midpoint of the signal segment is significantly greater than that at the other moments, so the vibration source with the greatest
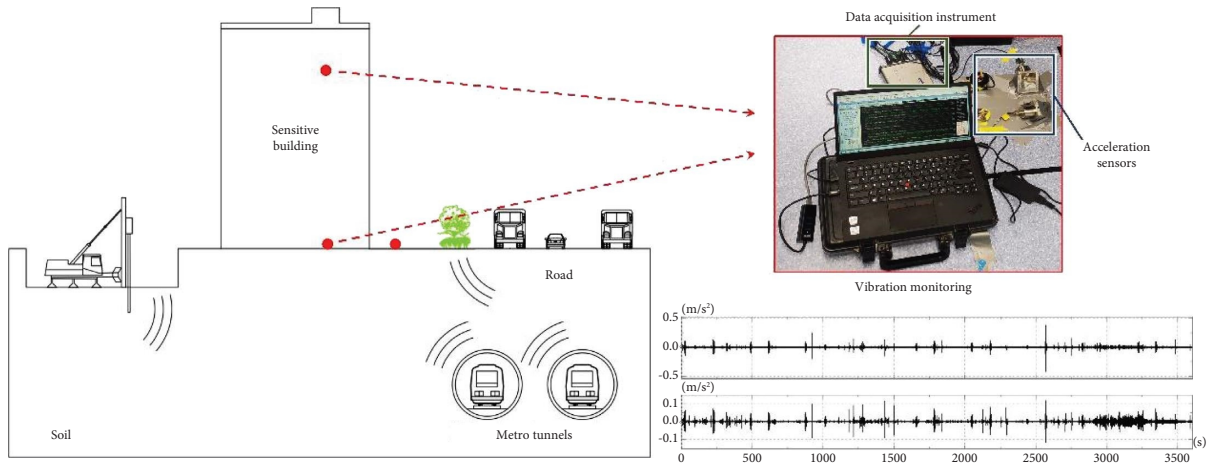
Figure 4: The monitoring of external vibration in the sensitive buildings.
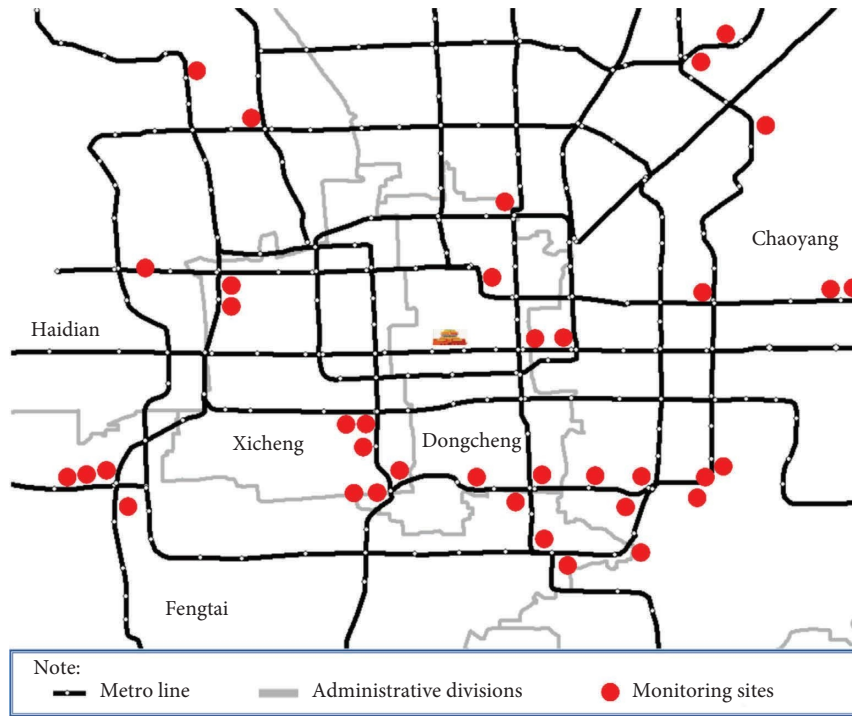

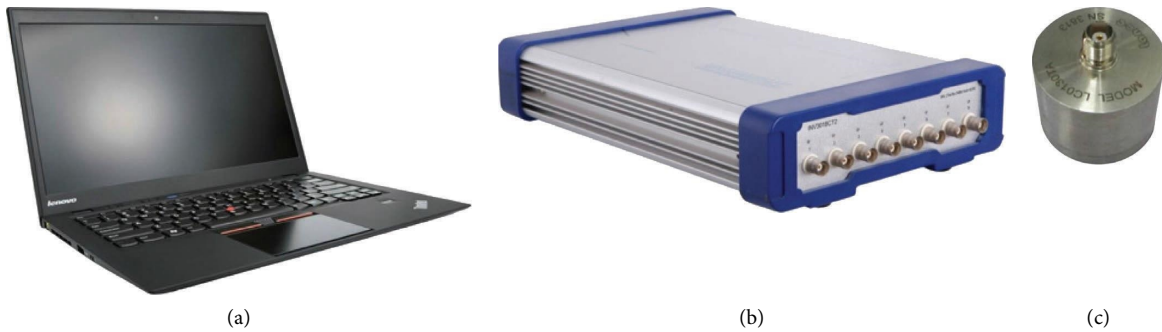Figure 5: The location of measurement sites of train-induced external vibration.

Figure 6: The equipment used in the external vibration monitoring: (a) computer; (b) data acquisition device; (c) acceleration sensor.
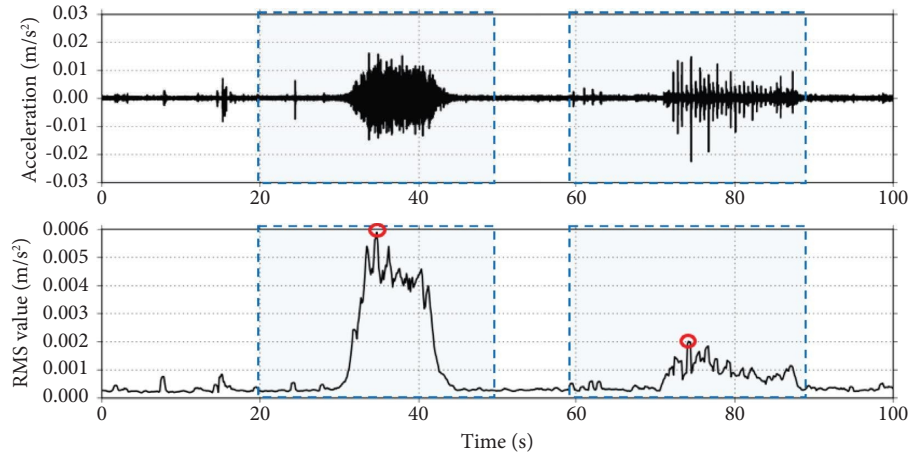
FIGURE 7: Illustration of the segmentation of vibration signals.

magnitude at the midpoint of the signal segment is used as the final label for the signal segment.

To obtain the most suitable time-frequency analysis technique for the monitored vibration signal, the effect of both HHT and CWT with different basis functions was investigated. Specifically, based on the sampling frequency of 512 Hz for the original monitoring signals, the analysis frequency for the time-frequency analysis was chosen to be 256 Hz, which is also common in external vibration analysis [58]. Besides, to increase the efficiency of the subsequent deep learning model, the scale of the time-frequency spectrum of each signal was transformed to 64×64. The results of the time-frequency analysis are shown in Figure 8. It can be seen that the four different vibration signals show more significant differences in characteristics in the time-frequency diagram compared to the waveform diagram, illustrating the necessity of preprocessing the signals using time-frequency analysis. Specifically, the energy of metro train-induced vibration in the time-frequency spectrum is usually concentrated in the frequency band of about 50 Hz and higher, which results from dynamic excitations generated by the contact between the tracks and the rails. While the energy of road vehicles induced vibration is usually more scattered and appears in the wavelet spectrum in the frequency band of about 30 Hz and lower. In addition, the vibrations generated by construction activities are usually manifested in the form of impulse signals, which appear in the spectrum as a sudden occurrence of energy and often cover a relatively wide range of frequency bands.

In addition, comparing the time-frequency spectra obtained by different techniques is shown in Figure 8. It can be found that the resolution of the time-frequency spectra obtained using HHT is higher than that of CWT, which makes it more sensitive to the features of small scales in the signal. But it only shows the energy of those vibration signals with the largest magnitude, while neglecting vibrations with relatively smaller magnitudes; this conclusion is consistent with some existing studies [59, 60]. This leads to the inability of HHT to depict the overall properties of the train and vehicle-induced vibrations on the time scale. In the results

obtained using CWTs with different basis functions, the results for CWTs with mexh functions show the lowest resolution on the frequency scale, which makes it possible to confuse the actual frequency of the vibration energy. In contrast, CWTs with Morl and Cgau functions show the greatest characteristic differences between different vibration signals and are therefore considered to be the most suitable time-frequency analysis for external vibration. Furthermore, considering that the performance of the CWT technique with Morl functions has been proven by more studies related to vibration signals processing [61, 62], it is therefore used as the final tool for time-frequency analysis in the subsequent studies of this paper.

It should be noted that although CWT showed relatively superior performance in identifying externals in this study, which is consistent with the conclusions obtained in some engineering applications [63, 64], HHT has also been considered to be superior to CWT in some other engineering applications [59, 60]. Therefore, researchers need to be careful in choosing a time-frequency analysis technique, as each technique has its unique advantages and disadvantages.

4. Model Implementation and Performance Evaluation

4.1. Data Description. According to the method described in Section 3, a large amount of external vibration data measured by the authors' research team in Beijing were preprocessed and were used for the training and testing of the proposed model in this study, and the sources of the measured external vibration include railway, road traffic, construction activities, and others. Specifically, the available samples are randomly divided into training and test samples in a ratio of 3 : 1. The number of vibration samples from different sources is shown in Table 1.

Since existing research [65] shows that using time-frequency spectra as input is faster and more space-efficient in storing data compared to time-frequency graphs, wavelet spectra are used as the model input in this paper, and the size of the wavelet spectra is set to 64×64.
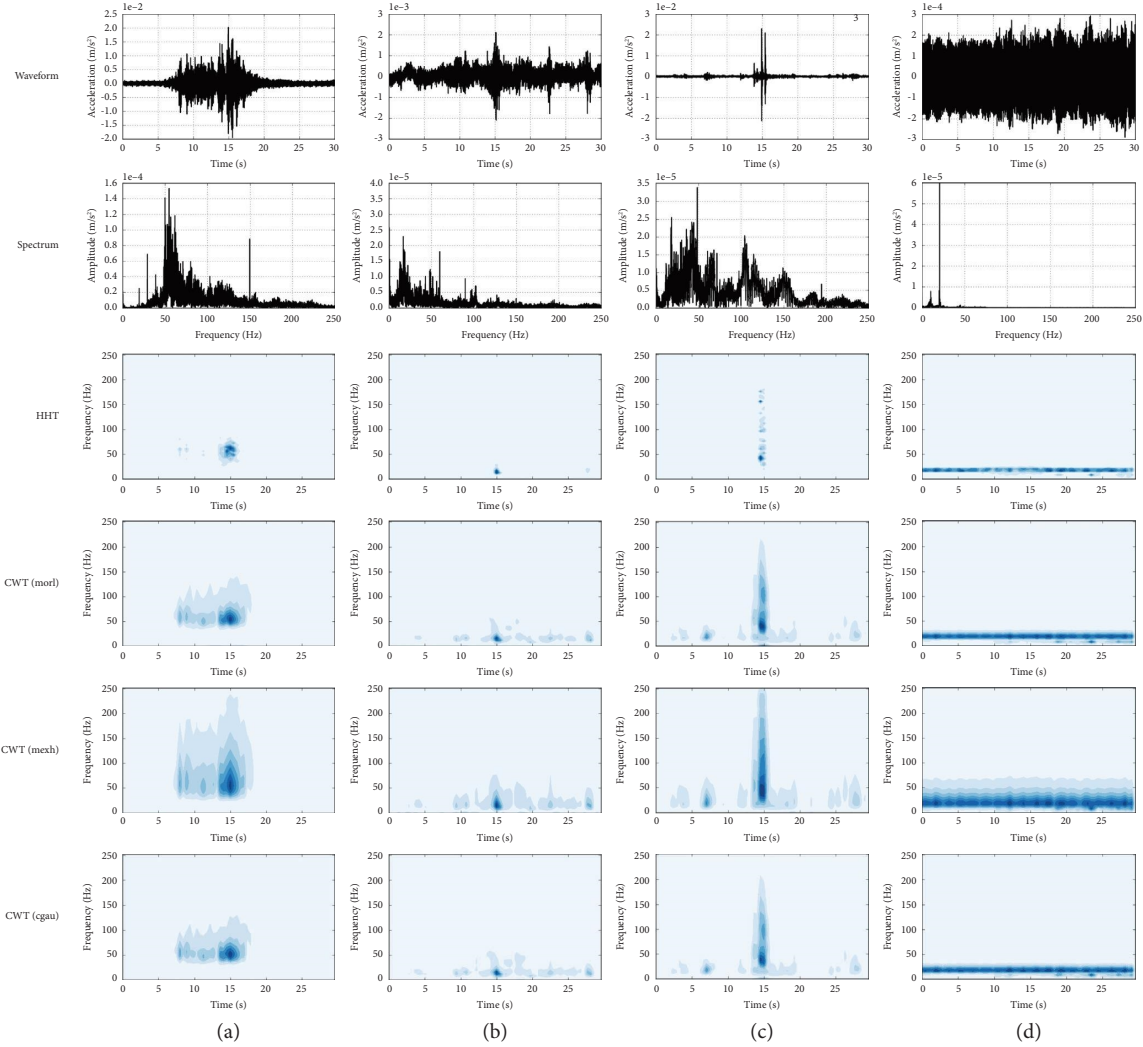
FIGURE 8: Time-frequency spectra of typical external vibrations: (a) metro train-induced vibration, (b) road vehicles-induced vibration, (c) construction activities-induced vibration, and (d) vibration from other sources.

4.2. Model Implementation. In this paper, three forms of hybrid models of CNN and LSTM and each submodel are built separately to investigate the most suitable deep learning model structure in external vibration identification. Considering that the hyperparameter settings in the models could significantly affect the model results, the random search method is used to optimize the hyperparameter combinations for each model. With reference to existing relevant research [26, 66], six parameters were set as hyperparameters to be optimized: the number of units in the LSTM, the number of CNN modules, the size and number of convolutional kernels in the CNN modules, the size of the pooling layer, and the number of FCNN units. Considering the capacity of computational devices, each hyperparameter was set at 2 or 3 levels and the sampling space of each hyperparameter is shown in Table 2. The number of iterations of random sampling for each model was set to 30 and the target indicator for optimization was the accuracy of the model on the test set.

Moreover, in model training, the dropout technique [45] is used after fully connected layers and LSTM layers in each model, and the dropout ratio was set as 50%. The loss

function is chosen as the mean squared error, and the optimizer is Adam [67] with a learning rate of 0.001. Considering the huge amount of computation in network training, mini-batch gradient descent was used as an optimizer to minimize losses and adjust weights in this experiment, this batch sampling strategy used less memory and was faster to train, and the batch samples were used in the training process and the batch size was set to 128, the training epoch was set to 100.

4.3. Model Performance and Discussion. The optimal hyperparameter combinations for each model obtained through random search are shown in Table 2, and the accuracy performance of each model with the optimal hyperparameter combinations is shown in Table 3. It can be found that the hybrid model of CNN+LSTM achieves the best performance among all the models with an accuracy of 97.7%, which is slightly higher than the other hybrid models. In addition, the CNN performed the best of the three submodels, with significantly higher accuracy than the

TABLE 1: Number of samples in each category.

Category of vibration source	Number of samples		
	Training set	Testing set	Total
Railway	2650	882	3532
Road traffic	2505	846	3351
Construction	303	91	394
Others	626	209	835
Total	6084	2028	8112

TABLE 2: The range and optimization results of the hyperparameters for each model.

Range and optimization results		LSTM block		CNN block			FCNN block
		Number of units	Number of blocks	Number of kernels	Kernel size	Pooling size	Number of units
	Range	32, 64, 128	2, 3	64, 128, 256	3, 5	2, 4	256, 512, 1024
Optimal results	CNN – LSTM	128	3	128	5	2	1024
	LSTM – CNN	64	2	256	5	2	1024
	CNN + LSTM	64	3	256	3	2	512
	CNN	—	2	256	3	2	1024
	LSTM	128	—	—	—	—	1024
	FCNN	—	—	—	—	—	1024

TABLE 3: The accuracy of each model for testing samples.

The maximum value of accuracy					
CNN – LSTM	LSTM – CNN	CNN + LSTM	CNN	LSTM	FCNN
97.3%	97.1%	97.7%	97.2%	96.0%	93.3%

LSTM and FCNN, and its accuracy was comparable with each of the hybrid models, indicating that the CNN can provide a stronger feature extraction capability in processing the time-frequency spectrum of vibration compared to the other submodels.

In addition, the accuracy of the hybrid model is not always higher than that of the submodel, specifically, the accuracy of the LSTM-CNN is slightly lower than that of the submodel CNN. The reason for this phenomenon might be that in this hybrid model, the front LSTM ignores some of the features in processing the input signal, thus decreasing the performance of the rear CNN in identifying the signal. This result indicates that the form of the hybrid model is very important and it needs to be carefully considered in relevant studies, an inappropriate hybrid model might achieve a lower accuracy performance than a submodel. In particular, the processing of the input signal in the hybrid model of CNN+LSTM is independent of each submodule, which could be the reason for the highest accuracy performance achieved by this form of hybrid model.

The identification performance of each model for each specific category is shown in Table 4, and the confusion matrix results obtained for each model on the testdataset are shown in Figure 9. It can be found that almost all models achieve the best performance in terms of accuracy and recall in identifying railway train-induced vibrations, and the worst performance in identifying construction activity-induced vibrations. The confusion in the identification results is mainly between vibration induced by construction activities and road traffic, and the number of confusion of

TABLE 4: Identification performance of models for each category.

Model	Indicator	Model performance for each category		
		Railway	Road traffic	Construction
CNN – LSTM	Precision	99.7%	97.5%	74.7%
	Recall	99.9%	96.5%	81.0%
	$F1$ score	0.998	0.970	0.777
LSTM – CNN	Precision	98.3%	99.1%	70.3%
	Recall	99.8%	96.4%	87.7%
	$F1$ score	0.990	0.977	0.780
CNN + LSTM	Precision	99.8%	99.2%	67.0%
	Recall	99.8%	95.9%	92.4%
	$F1$ score	0.998	0.975	0.777
CNN	Precision	98.9%	98.3%	72.5%
	Recall	99.8%	96.7%	82.5%
	$F1$ score	0.993	0.975	0.772
LSTM	Precision	99.7%	98.0%	46.2%
	Recall	99.7%	94.3%	76.4%
	$F1$ score	0.997	0.961	0.575
FCNN	Precision	98.8%	95.3%	22.0%
	Recall	99.9%	91.2%	39.2%
	$F1$ score	0.993	0.932	0.282

railway-induced vibration is relatively few. There could be two reasons for this phenomenon. First, the differences between the characteristics of train-induced vibration and those of other vibrations in the time-frequency spectrum are significant, so it is simpler to distinguish them in the identification. Whereas the differences between the vibration characteristics of construction activities and those of road

traffic are relatively small in the time-frequency spectrum, as they overlap in the frequency band below 50 Hz, which could lead to confusion between them by models. In addition, since the number of signals collected in this study for construction activity-induced vibration is less than the other two, this could also lead to the worst identification performance for construction activity-induced vibration.

5. Case Study

In this section, a real-world case of external vibration monitoring is presented, and the proposed method is used for the identification of the sources of vibration recorded from the monitoring so that the vibration from each source can be evaluated, and the results of the evaluation and the performance of the model are demonstrated and discussed.

5.1. Background. In this case, external vibration monitoring and evaluation were requested for a laboratory. The laboratory is equipped with several precision instruments, and excessive external vibrations could affect the normal operation of these precision instruments, such as the imaging quality of microscopes. Therefore, vibration from various external sources in the laboratory needs to be evaluated separately to provide evidence for subsequent targeted vibration mitigation design. In this case, the generic vibration criteria (VC) [68] standard was used to evaluate the level of vibration impact on the laboratory, and this standard proposes seven levels of vibration magnitude to quantify the impact of vibration on the instrumentations.

The location of the building is shown in Figure 10, in which the laboratory is labeled as Building A, and the external vibration sources outside the building include road traffic and underground railway. Specifically, the horizontal distance between the metro tunnel and the laboratory is 31 m, while the closest distance between the road and the laboratory is only 13 m.

To evaluate the impact of external vibrations on the laboratory accurately, external vibration monitoring was carried out in the laboratory. A vibration monitoring point was installed on the ground floor of the laboratory and monitoring was conducted continuously for up to 24 hours; the location of the measurement point is shown in Figure 10. After obtaining the results of the vibration monitoring, the method proposed in this paper was utilized to determine the source of the vibration segments and then an accurate assessment of the external vibration from each source was carried out.

5.2. External Vibration Source Identification. Following the signal preprocessing method described in Section 2.2, the monitored vibration signals in this study were segmented and analyzed using CWT, and the vibration source of each signal segment was identified by the optimal CNN+LSTM model trained and tested in the above section.

The performance of the proposed model in this case for vibration source identification is shown in Table 5 and Figure 11; it can be found that the $F1$ scores of the model for

the identification of vibrations induced by railway and road traffic are higher than 0.95, which is similar to the performance achieved by the model on the testdataset in the previous section. This further verifies the feasibility of the proposed method in this paper.

To further investigate the reasons for model misidentification on some samples, several typical samples misidentified by the model are shown in Figure 12. It can be seen that there could be two possible causes of model misidentification. The first is because some of the vibrations exhibit features on the time-frequency spectrum that are very similar to those of vibration signals from other categories, causing the model to misidentify them, and such a case is shown in Figure 12(a). To reduce this type of misidentification, it might be effective to add some vibration signals that are easily confused as training samples.

In addition, some signal segments can record vibration signals from multiple sources at the same time, and when experts label these samples, they usually use the category of the vibration with the largest magnitude as the label. However, when processing such signals, the model might only focus on the characteristics of vibration signals with smaller magnitudes, thus causing misidentification, and such cases are shown in Figures 12(b)–12(d). To reduce such misidentifications, it might be effective to add some quantitative metrics describing the vibration magnitude as input to the model, which needs to be further investigated.

5.3. Evaluation of External Vibration. The 5%, 25%, 50%, 75%, and 95% percentiles of the identified external vibration from the two sources were calculated using statistical analysis, the comparison of the results with the limit values proposed by standard VC for each level is shown in Figure 13. In this case, the vibration impact generated by both railway and road traffic on the sensitive instruments in the laboratory is severe, the vibration impact caused by the railway is mainly concentrated in the frequency band near 50 Hz, while the vibration impact caused by road traffic is mainly concentrated in the frequency band near 20 Hz.

Although the median value of vibration caused by railway and road traffic is less than the limit value of VC-D level, the 5th percentile of train-induced vibration is greater than the limit value of VC-A level, while the 25th percentile is greater than the limit value of VC-C level. The 5th percentile of vibration caused by road traffic is greater than the limit value of the VC-B level, and the 25th percentile is greater than the limit value of the VC-D level. This indicates that the impact of vibration induced by railway on the laboratory could potentially be more significant than that induced by road traffic. These detailed vibration evaluation results could be beneficial for the subsequent vibration mitigation design of the laboratory.

In addition, it also can be seen that the evaluation results obtained from the identified vibration of the two sources using statistical analysis are very close to the results obtained from the actual vibration of the two sources, which further demonstrates the feasibility of the method proposed in this paper.
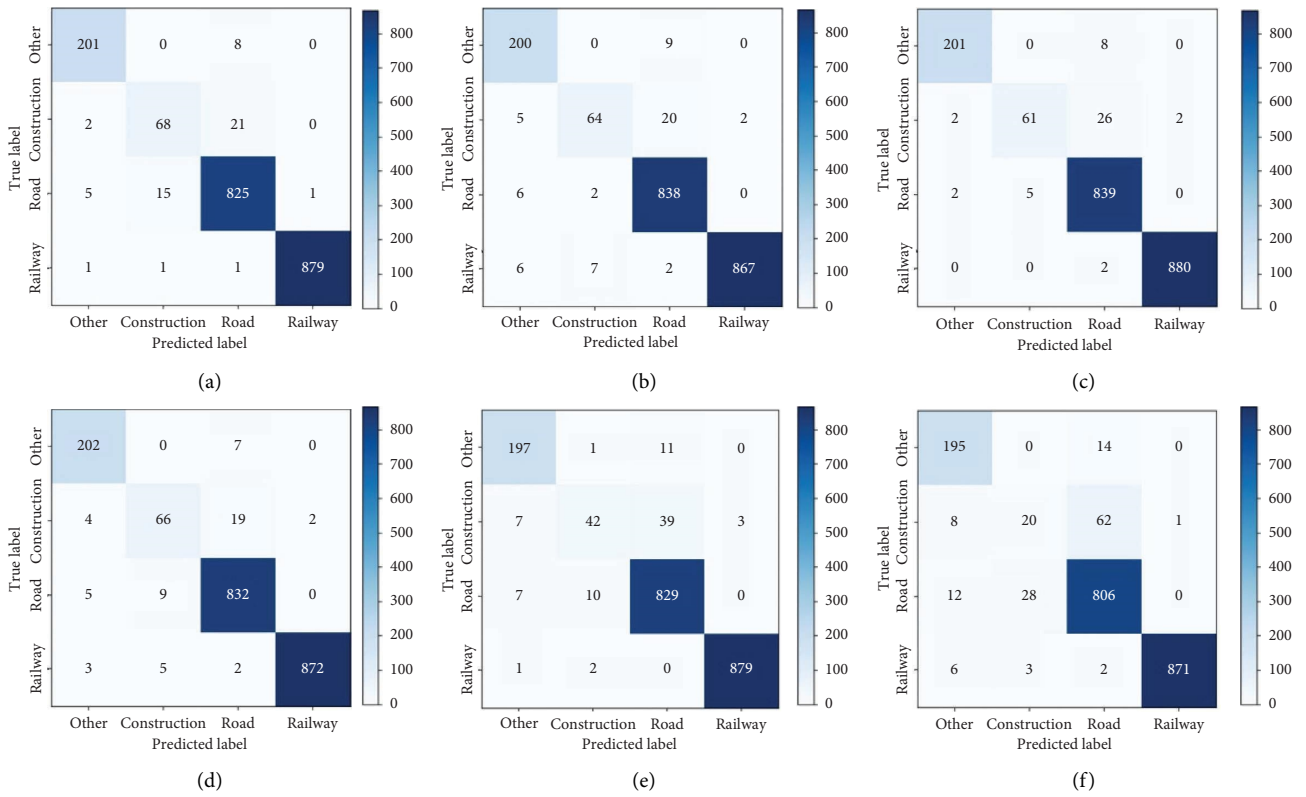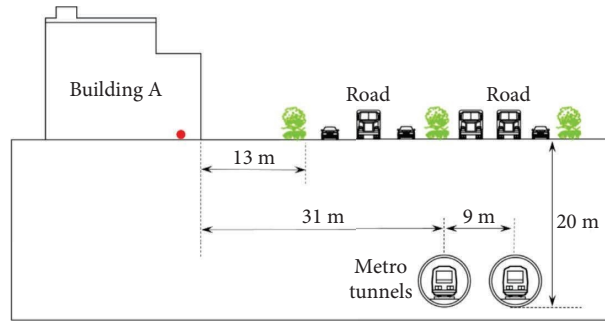
FIGURE 9: Confusion matrix of testing results of each model: (a) CNN – LSTM, (b) LSTM – CNN, (c) CNN + LSTM, (d) CNN, (e) LSTM, and (f) FCNN.

Notes: ● Environmental vibration monitoring point

FIGURE 10: Location of the building being monitored for vibration and the surrounding vibration sources.

TABLE 5: Model performance for each category in the case study.

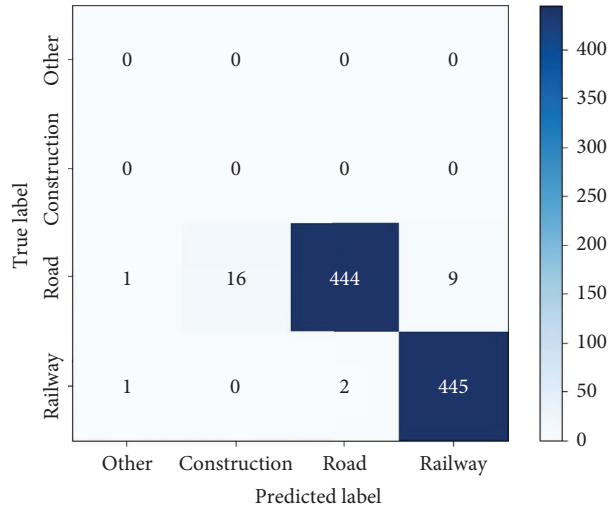| Indicator	Model performance for each category	
	Railway	Road traffic
Precision	99.3%	94.5%
Recall	98.0%	99.6%
$F1$ score	0.987	0.969

Figure 11: Confusion matrix of model identification results in the case study.
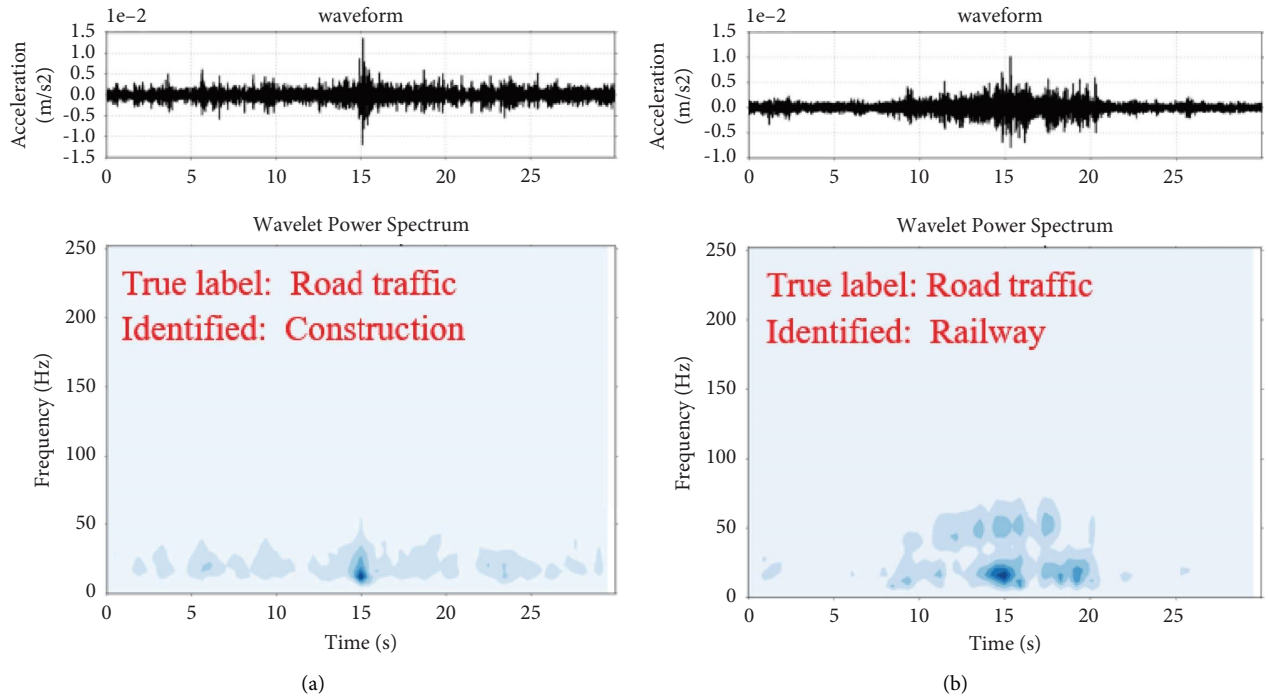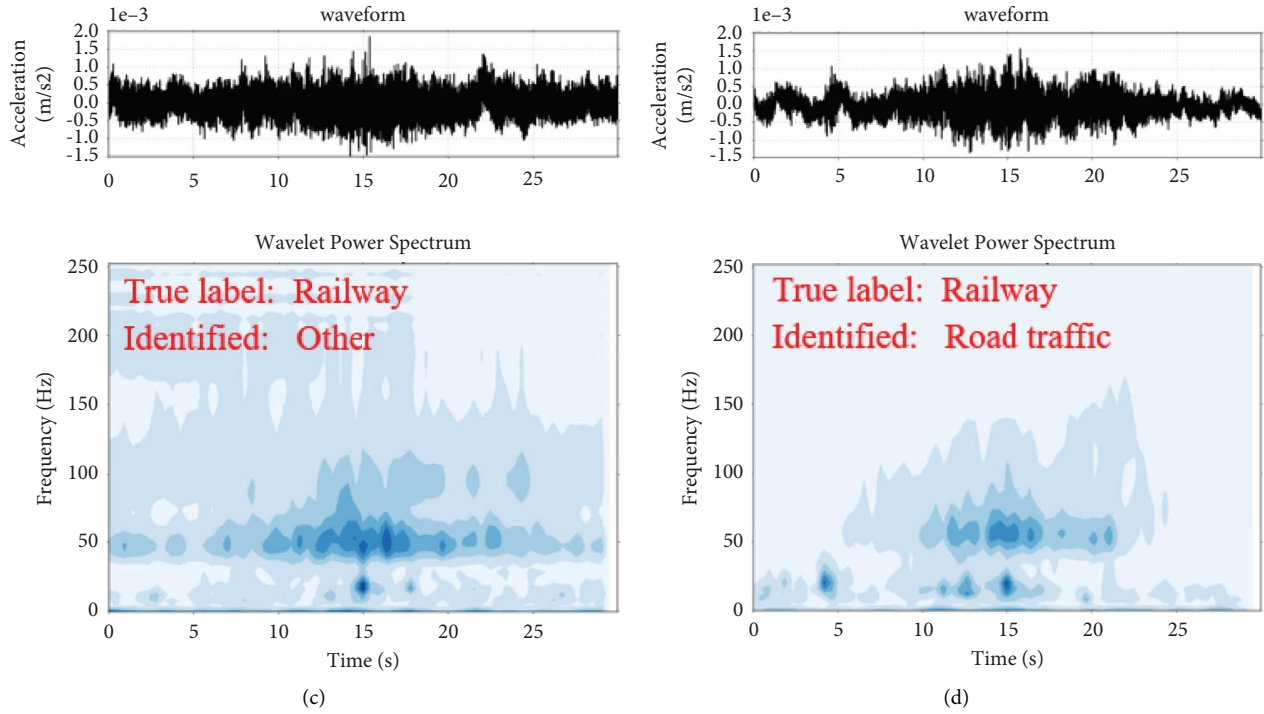

(a)

(b)

Figure 12: Continued.

(c)

(d)

FIGURE 12: Typical samples incorrectly identified by the model.
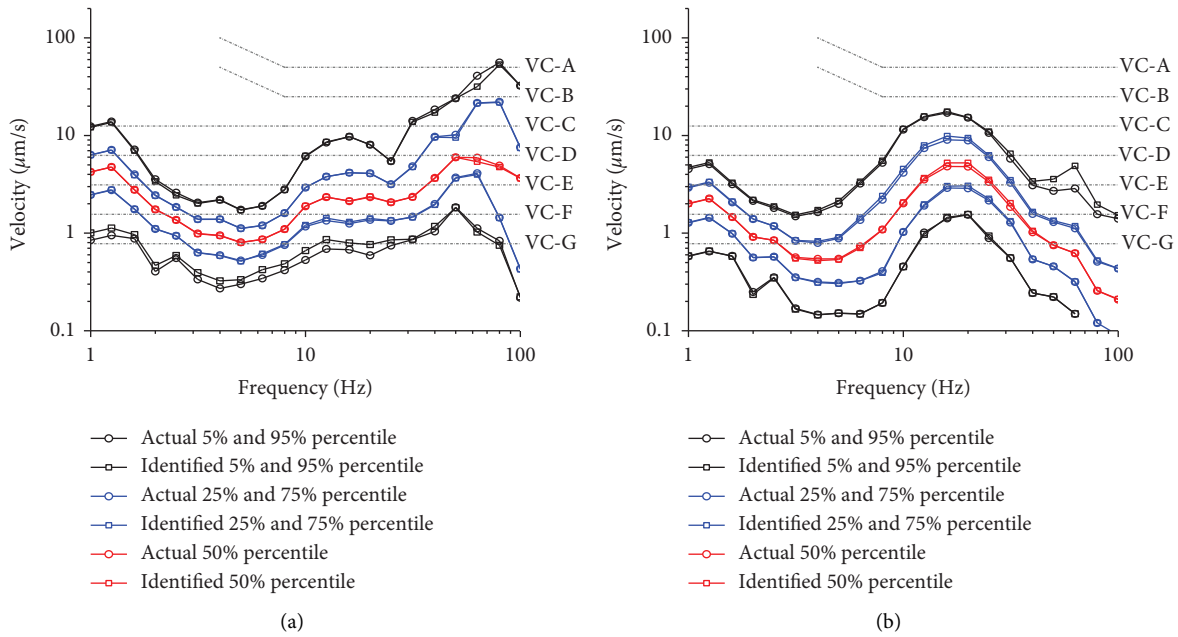

(a)

(b)

FIGURE 13: Evaluation of vibration caused by (a) railway and (b) road traffic on sensitive buildings using statistical analysis.

6. Conclusion

To improve the efficiency of external vibration monitoring and evaluation in sensitive buildings, this paper proposes a deep learning-based external vibration source identification method, which can automatically identify vibration sources in the monitoring process of external vibration, thus facilitating accurate probabilistic analysis of external vibration impact from various sources at sensitive buildings and subsequent targeted design of vibration mitigation measures.

Specifically, hybrid deep learning models with multiple structures of CNN and LSTM and each submodel were built separately. These models are trained and tested using extensive external vibration data measured in Beijing, which are transformed into time-frequency spectra using CWT.

After testing the performance of various models, the hybrid model of CNN+LSTM was proved to be the optimal model for external vibration source identification, with an accuracy of over 97%. In addition, the submodel CNN also exhibits strong performance in the classification of CWT spectra of signals, and its accuracy is only slightly lower than that of the optimal hybrid model. However, those vibration signal segments that record multiple sources are potentially misidentified by the deep learning model, which is the difficulty of this work and needs to be investigated further.

Finally, this paper presents a real-world case study in which a laboratory with sensitive instruments was monitored for external vibration. The sources of the monitored vibration signal segments were identified using the model of CNN+LSTM proposed in this paper, and the impact of the vibration from each source on sensitive buildings was evaluated using statistical analysis, which further validates the necessity and feasibility of the proposed method in engineering application.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L. Xu and M. Ma, "Analytical solution of ground-borne vibration due to a spatially periodic harmonic moving load in a tunnel embedded in layered soil," *Journal of Zhejiang University - Science*, 2023.

[2] W. Liu, C. Li, L. Ma, and L. Du, "A frequency-domain formulation for predicting ground-borne vibration induced by underground train on curved track," *Journal of Sound and Vibration*, vol. 549, Article ID 117578, 2023.

[3] Y. Li, H. Li, G. Zhang, and S. Kaewunruen, "Nonlinear responses of longitudinally coupled slab tracks exposed to extreme heat waves," *Engineering Structures*, vol. 281, Article ID 115789, 2023.

[4] R. Liang, W. Liu, W. Li, and Z. Wu, "A traffic noise source identification method for buildings adjacent to multiple transport infrastructures based on deep learning," *Building and Environment*, vol. 211, Article ID 108764, 2022.

[5] O. Tavasoli and M. Ghazavi, "Wave propagation and ground vibrations due to non-uniformcross-sections piles driving," *Computers and Geotechnics*, vol. 104, pp. 13–21, 2018.

[6] K. Wei, Z. Zhao, X. Du, H. Li, and P. Wang, "A theoretical study on the train-induced vibrations of a semi-activemagneto-rheologicalsteel-spring floating slab track," *Construction and Building Materials*, vol. 204, pp. 703–715, 2019.

[7] H. Fu, Y. Yang, and S. Kaewunruen, "Multi-hazard effects of crosswinds on cascading failures of conventional and interspersed railway tracks exposed to ballast washaway and moving train loads," *Sensors*, vol. 23, no. 4, 2023.

[8] M. Ma, B. Jiang, W. Liu, and K. Liu, "Control of metro train-induced vibrations in a laboratory using periodic piles," *Sustainability*, vol. 12, no. 14, 2020.

[9] W. Li and W. Liu, "A novel formulation for transfer path identification and vibration prediction in the over-track building induced by trains," *Environmental Science and Pollution Research*, 2023.

[10] D. Ulgen, O. Ertugrul, and M. Ozkan, "Measurement of ground borne vibrations for foundation design and vibration isolation of a high-precision instrument," *Measurement*, vol. 93, pp. 385–396, 2016.

[11] W. Liu, Z. Wu, C. Li, and L. Xu, "Prediction of ground-borne vibration induced by a moving underground train based on excitation experiments," *Journal of Sound and Vibration*, vol. 523, Article ID 116728, 2022.

[12] M. Ma, M. Li, X. Qu, and H. Zhang, "Effect of passing metro trains on uncertainty of vibration source intensity: monitoring tests," *Measurement*, vol. 193, Article ID 110992, 2022.

[13] X. Qu, D. Thompson, M. Ma, M. Li, and E. Ntotsios, "Sources of variability in metro train-induced vibration," *Proceedings of the Institution of Mechanical Engineers - Part F: Journal of Rail and Rapid Transit*, Article ID 095440972211184, 2022.

[14] Y. Pan and L. Zhang, "Roles of artificial intelligence in construction engineering and management: a critical review and future trends," *Automation in Construction*, vol. 122, Article ID 103517, 2021.

[15] W. Liu, R. Liang, H. Zhang, Z. Wu, and B. Jiang, "Deep learning based identification and uncertainty analysis of metro train induced ground-borne vibration," *Mechanical Systems and Signal Processing*, vol. 189, Article ID 110062, 2023.

[16] R. Liang, W. Liu, C. Li, W. Li, and Z. Wu, "A novel efficient probabilistic prediction approach for train-induced ground vibrations based on transfer learning," *Journal of Vibration and Control*, Article ID 107754632211487, 2023.

[17] R. Liang, W. Liu, M. Ma, and W. Liu, "An efficient model for predicting the train-inducedground-borne vibration and uncertainty quantification based on Bayesian neural network," *Journal of Sound and Vibration*, vol. 495, Article ID 115908, 2021.

[18] C. Li and W. Liu, "Probabilistic prediction of metro induced ground-borne vibration and its accuracy evaluation," *Soil Dynamics and Earthquake Engineering*, vol. 141, Article ID 106509, 2021.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 60, 2012.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] O. Alazzawi and D. Wang, "A novel structural damage identification method based on the acceleration responses

under ambient vibration and an optimized deep residual algorithm," *Structural Health Monitoring*, vol. 21, Article ID 14759217211065009, 2022.

[22] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mechanical Systems and Signal Processing*, vol. 107, pp. 241–265, 2018.

[23] T. Han, T. Zhou, Y. Xiang, and D. Jiang, "Cross-machine intelligent fault diagnosis of gearbox based on deep learning and parameter transfer," *Structural Control and Health Monitoring*, vol. 29, no. 3, Article ID e2898, 2022.

[24] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, and D. J. Inman, "A review of vibration-based damage detection in civil structures: from traditional methods to Machine Learning and Deep Learning applications," *Mechanical Systems and Signal Processing*, vol. 147, Article ID 107077, 2021.

[25] B. K. Oh, Y. Park, and H. S. Park, "Seismic response prediction method for building structures using convolutional neural network," *Structural Control and Health Monitoring*, vol. 27, no. 5, Article ID e2519, 2020.

[26] R. Yang, S. K. Singh, M. Tavakkoli, N. Amiri, M. A. Karami, and R. Rai, "Continuous video stream pixel sensor: a CNN-LSTM based deep learning approach for mode shape prediction," *Structural Control and Health Monitoring*, vol. 29, no. 3, Article ID e2892, 2022.

[27] U. Javed, K. Ijaz, M. Jawad et al., "A novel short receptive field based dilated causal convolutional network integrated with bidirectional LSTM for short-term load forecasting," *Expert Systems with Applications*, vol. 205, Article ID 117689, 2022.

[28] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM model for short-term individual household load forecasting," *IEEE Access*, vol. 8, Article ID 180544, 2020.

[29] R. Barzegar, M. T. Aalami, and J. Adamowski, "Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model," *Stochastic Environmental Research and Risk Assessment*, vol. 34, no. 2, pp. 415–433, 2020.

[30] U. Pak, C. Kim, U. Ryu, K. Sok, and S. Pak, "A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction," *Air Quality, Atmosphere and Health*, vol. 11, no. 8, pp. 883–895, 2018.

[31] A. Vidal and W. Kristjanpoller, "Gold volatility prediction using a CNN-LSTM approach," *Expert Systems with Applications*, vol. 157, Article ID 113481, 2020.

[32] H. Rezaei, H. Faaljou, and G. Mansourfar, "Stock price prediction using deep learning and frequency decomposition," *Expert Systems with Applications*, vol. 169, Article ID 114332, 2021.

[33] K. Nandhini and G. Tamilpavai, "Hybrid CNN-LSTM and modified wild horse herd Model-based prediction of genome sequences for genetic disorders," *Biomedical Signal Processing and Control*, vol. 78, Article ID 103840, 2022.

[34] S. Bagherzadeh, M. S. Shahabi, and A. Shalbaf, "Detection of schizophrenia using hybrid of deep learning and brain effective connectivity image from electroencephalogram signal," *Computers in Biology and Medicine*, vol. 146, Article ID 105570, 2022.

[35] G. Petmezas, K. Haris, L. Stefanopoulos et al., "Automated atrial fibrillation detection using a hybrid CNN-LSTM network on imbalanced ECG datasets," *Biomedical Signal Processing and Control*, vol. 63, Article ID 102194, 2021.

[36] P. Liang, C. Deng, J. Wu, Z. Yang, J. Zhu, and Z. Zhang, "Compound fault diagnosis of gearboxes via multi-label convolutional neural network and wavelet transform," *Computers in Industry*, vol. 113, Article ID 103132, 2019.

[37] N. E. Huang, Z. Shen, S. R. Long et al., "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp. 903–995, 1998.

[38] N. E. Huang and Z. Wu, "A review on Hilbert-Huang transform: method and its applications to geophysical studies," *Reviews of Geophysics*, vol. 46, no. 2, 2008.

[39] D. Čeponis and N. Goranin, "Investigation of dual-flow deep learning models LSTM-FCN and GRU-FCN efficiency against single-flow CNN models for the host-based intrusion and malware detection task on univariate times series data," *Applied Sciences*, vol. 10, no. 7, 2020.

[40] M. J. Hasan, M. M. Islam, and J.-M. Kim, "Multi-sensorfusion-basedtime-frequency imaging and transfer learning for spherical tank crack diagnosis under variable pressure conditions," *Measurement*, vol. 168, Article ID 108478, 2021.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.

[42] C. Szegedy, W. Liu, Y. Jia et al., "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, Boston, MA, USA, June 2015.

[43] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, Haifa, Israel, June 2010.

[44] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the International conference on machine learning, PMLR*, pp. 448–456, Lille France, July 2015.

[45] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[46] B. Zhao, C. Cheng, Z. Peng, X. Dong, and G. Meng, "Detecting the early damages in structures with nonlinear output frequency response functions and the CNN-LSTM model," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9557–9567, 2020.

[47] B. Wang, S. Liu, B. Wang, W. Wu, J. Wang, and D. Shen, "Multi-step ahead short-term predictions of storm surge level using CNN and LSTM network," *Acta Oceanologica Sinica*, vol. 40, no. 11, pp. 104–118, 2021.

[48] X. Ye, H. Ren, J. Nie et al., "Simultaneous estimation of land surface and atmospheric parameters from thermal hyperspectral data using a LSTM–CNN combined deep neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[49] L. Fu, Q. Tang, P. Gao, J. Xin, and J. Zhou, "Damage identification of long-span bridges using the hybrid of convolutional neural network and long short-term memory network," *Algorithms*, vol. 14, no. 6, 2021.

[50] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

[51] J. Bergstra and Y. Bengio, "Random search for hyperparameter optimization," *Journal of Machine Learning Research*, vol. 13, 2012.

[52] T.-P. Nguyen, C.-T. Yeh, M.-Y. Cho, C.-L. Chang, and M.-J. Chen, "Convolutional neural network bidirectional long

short-term memory to online classify the distribution insulator leakage currents," *Electric Power Systems Research*, vol. 208, Article ID 107923, 2022.

[53] I. Priyadarshini and C. Cotton, "A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis," *The Journal of Supercomputing*, vol. 77, no. 12, Article ID 13911, 2021.

[54] M. Hosseini, D. Bani-Hani, and S. S. Lam, "Leukocytes image classification using optimized convolutional neural networks," *Expert Systems with Applications*, vol. 205, Article ID 117672, 2022.

[55] L. Ducarne, D. Ainalis, and G. Kouroussis, "Assessing the ground vibrations produced by a heavy vehicle traversing a traffic obstacle," *Science of the Total Environment*, vol. 612, pp. 1568–1576, 2018.

[56] M. Stolarik, M. Pinka, and J. Nedoma, "Ground-borne vibration due to construction works with respect to brownfield areas," *Applied Sciences*, vol. 9, no. 18, 2019.

[57] National Environmental Protection Agency, *GB 10071-88: Measurement Method of Environmental Vibration in Urban Area*, National Environmental Protection Agency, Beijing, China, 1988.

[58] R. Liang, D. Ding, W. Liu, F. Sun, and Y. Cheng, "Experimental study of the source and transmission characteristics of train-induced vibration in the over-track building in a metro depot," *Journal of Vibration and Control*, Article ID 107754632110701, 2022.

[59] A. Biswas, "Scale–location specific soil spatial variability: a comparison of continuous wavelet transform and Hilbert–Huang transform," *Catena*, vol. 160, pp. 24–31, 2018.

[60] A. Y. Ayenu-Prah and N. O. Attoh-Okine, "Comparative study of Hilbert–Huang transform, Fourier transform and wavelet transform in pavement profile analysis," *Vehicle System Dynamics*, vol. 47, no. 4, pp. 437–456, 2009.

[61] M.-Q. Tran, M.-K. Liu, Q.-V. Tran, and T.-K. Nguyen, "Effective Fault diagnosis based on wavelet and convolutional attention neural network for induction motors," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.

[62] R. C. dos Santos, A. P. C. Larocca, J. O. de Araújo Neto, A. C. B. Barbosa, and J. V. M. Oliveira, "Detection of a curved bridge deck vibration using robotic total stations for structural health monitoring," *Journal of Civil Structural Health Monitoring*, vol. 9, no. 1, pp. 63–76, 2019.

[63] M. Stork and V. Vancura, "Wavelet and Hilbert-Huang transform used in cardiology," in *Proceedings of the 2014 6th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pp. 61–66, Bucharest, Romania, October 2014.

[64] A. Bagheri, K. Li, and P. Rizzo, "Reference-free damage detection by means of wavelet transform and empirical mode decomposition applied to Lamb waves," *Journal of Intelligent Material Systems and Structures*, vol. 24, no. 2, pp. 194–208, 2013.

[65] X. Lu, W. Liao, W. Huang, Y. Xu, and X. Chen, "An improved linear quadratic regulator control method through convolutional neural network–based vibration identification," *Journal of Vibration and Control*, vol. 27, no. 7-8, pp. 839–853, 2021.

[66] S. Ma, L. Gao, X. Liu, and J. Lin, "Deep learning for track quality evaluation of high-speed railway based on vehicle-body vibration prediction," *IEEE Access*, vol. 7, Article ID 185099, 2019.

[67] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, https://arxiv.org/abs/1412.6980.

[68] H. Amick, "On generic vibration criteria for advanced technology facilities: with a tutorial on vibration data representation," *Journal of the IEST*, vol. 40, no. 5, pp. 35–44, 1997.