Research Article

Structural damage identification method based on Swin Transformer and continuous wavelet transform

Jingzhou Xin[1] (ID)**, Guangjiong Tao[1], Qizhi Tang[1]** (ID)**, Fei Zou[2], Chenglong Xiang[2]**

[1]State Key Laboratory of Mountain Bridge and Tunnel Engineering, Chongqing Jiaotong University, Chongqing 400074, China.
[2]Department of Road and Bridge Engineering, Guizhou Communications Polytechnic, Guiyang 551400, Guizhou, China.

Correspondence to: Dr. Qizhi Tang, State Key Laboratory of Mountain Bridge and Tunnel Engineering, Chongqing Jiaotong University, No.66 Xuefu Street, Nanan District, Chongqing 400074, China. E-mail: tangqizhi1015@163.com

Abstract

The accuracy improvement of deep learning-based damage identification methods has always been pursued. To this end, this study proposes a novel damage identification method using Swin Transformer and continuous wavelet transform (CWT). Specifically, the original structural vibration data is first transferred to a time-frequency diagram by CWT, thereby capturing the characteristic information of structural damage. Secondly, the Swin Transformer is applied to learn the two-dimensional time-frequency diagram layer by layer and extract the damage information, by which the damage identification is achieved. Then, the identification accuracy of the proposed method is analyzed under various sample lengths and different levels of environmental noise to validate the robustness of this approach. Finally, the practicality of this method is verified through laboratory test. The results show the proposed method can effectively recognize the damage and achieve excellent accuracy even under noise interference. Its accuracy reaches 99.6% and 99.0% under single damage and multiple damage scenarios, respectively.

Keywords: Artificial intelligence, deep learning, damage identification, Swin Transformer, continuous wavelet transform

INTRODUCTION

Engineering structures are prone to damage in complex service environments, which can imperil the

structural safety. Consequently, it is of paramount scientific significance to identify both the location and extent of damage using monitoring data in engineering structures[1].

Vibration-based damage identification methods[2,3] have been widely studied by many scholars in recent years, which can be divided into parameter-, machine learning-, and deep learning-based methods. The parameter-based approaches realize damage identification by extracting damage-sensitive features from the vibration mode, modal strain energy and other data. For example, Daneshvar *et al.* developed a new modal strain energy sensitivity function to improve damage detectability, and successfully localize and quantify damage under incompletely noisy modal data[4]. Pooya *et al.* used the absolute difference of modal strain energy coefficients as an indicator of damage location and applied the relationship between modal strain energy and modal kinetic energy to identify the damage of beam[5]. An *et al.* proposed a damage identification method for semi-rigid joints in frame structures based on additional virtual mass, which utilizes natural frequencies to identify the location and extent of damage[6]. Although parameter-based methods can be effective in identifying damage, they may not be sensitive enough to detect local damage and may have certain limitations.

In contrast, machine learning-based methods achieve recognition by learning labeled information and extracting features from it, such as random forest[7], artificial neural network (ANN)[8], and support vector machine (SVM)[9]. For example, Ren *et al.* proposed a method to identify damaged cables in cable-stayed bridges from the bridge deck bending strain response using SVM[10]. Farias *et al.* proposed the multi-particle collision algorithm to design an optimal ANN architecture for detecting and locating damage in plate structures[11]. Although machine learning-based methods can autonomously learn damage characteristics, they may not produce accurate results for complex nonlinear data.

Compared to machine learning, deep learning is better equipped to process large amounts of complex nonlinear data. The mainstream deep learning techniques currently include convolutional neural networks (CNNs)[12], deep neural networks (DNNs)[13], recurrent neural networks (RNNs)[14], and long short-term memory (LSTM)[15], which are widely used in damage recognition. For example, Fu *et al.* combined CNNs and LSTM to predict the location and severity of the bridge damage[16]. Fernandez-Navamuel *et al.* incorporated supervised learning and DNN to accurately identify damage types under different environments and operating conditions[17]. Sony *et al.* proposed a method for damage detection in full-scale bridges using a windowed one-dimensional CNN, which is effective for different types of damage[18]. Compared to the other two methods, deep learning can handle a large amount of nonlinear data and is more sensitive to the characteristic information of local damage. Therefore, deep learning-based methods are more advantageous for identifying damage in bridge structures. However, further endeavors are still needed to increase the accuracy of damage identification.

Swin Transformer[19] has gained popularity in various fields due to its exceptional image processing capabilities. For example, Üzen *et al.* utilized Swin Transformer to detect surface defects at the pixel level[20]. Xu *et al.* proposed a self-integrated Swin Transformer network structure, which combines the features of different layers of the Swin Transformer network and removes noisy points present in a single layer, thereby enhancing the retrieval performance[21]. The Swin Transformer is commonly utilized in various engineering fields. It has demonstrated excellent learning capabilities and recognition effects in processing image data. This can provide new ideas for improving the accuracy of structural damage identification.

In view of the shortcomings of existing research, this paper proposes a structural damage identification method based on Swin Transformer and continuous wavelet transform (CWT). First, the original data is

transferred to the image feature space through CWT. This process transforms the data into a two-dimensional RGB image; Secondly, the Swin Transformer model is applied to process the nonlinear and non-stationary signals. This allows for learning from time and space features to complete model training. The trained model is then used to accept the test set and perform feature matching to identify damage; Then, the robustness of the proposed method is verified under varying sample lengths and environmental noise interference; Finally, the practicality of this method is verified by experimental test. The main contributions of this paper are summarized below:

(1) The damage identification method combining Swin Transformer and CWT is proposed to improve the identification accuracy. The results demonstrate the superior performance of the proposed method compared to several other models, indicating that the Swin Transformer model may be an effective approach.

(2) A comprehensive parameter analysis is performed to examine the practicability of the proposed method. The influence of various sampling frequencies and sample lengths on damage identification is investigated, as well as the effect on model recognition under noise interference. By this analysis, the practicability of the proposed method can be verified, and the guidance under similar scenarios may be produced.

METHODS
Swin Transformer
Figure 1 illustrates the overall architecture of Swin Transformer[19]. In "Stage 1", the time-frequency diagram, with a size of [$H \times W \times 3$], is divided into non-overlapping patches with a feature dimension of $4 \times 4 \times 3 = 48$; then, the non-overlapping image patches are projected to an arbitrary dimension C through a linear embedding layer. The Swin Transformer model is applied to these patch markers, producing feature maps of size [$H/4$, $W/4$, C]; In "Stage 2", the number of tokens is reduced by patch-merging layers to create a hierarchical representation. The patch merging layer is comparable to the pooling layer in CNNs. Downsampling is performed before the start of each stage to reduce the resolution, and the number of channels is adjusted to form a hierarchical structure. The patch merging layer concatenates the features of each group of 2×2 adjacent patches and sets the output dimension to $2C$. The transformation of features is carried out using two consecutive Swin Transformer modules, resulting in a feature map of size [$H/8$, $W/8$, $2C$]; In "Stage 3", the procedures akin to those delineated in "Stage 2" are reiterated, culminating in the generation of a feature map sized [$H/16$, $W/16$, $4C$]; In "Stage 4", the steps of "Stage 2" are repeated, resulting in the output of a feature map with a size of [$H/32$, $W/32$, $8C$].

Swin Transformer block
The Swin Transformer block[19] replaces the standard multi-head self-attention (MSA) module in the traditional Transformer with a module based on a shift window. This change improves the efficiency of the model while maintaining the same level of accuracy. Figure 2 shows the Swin Transformer model, comprising a window-based MSA module (W-MSA), a shift window-based MSA module (SW-MSA), and a double-layer Multilayer Perceptron (MLP). The MLP layer is a common fully connected neural network layer. It consists of multiple fully connected layers and nonlinear activation functions. Each MSA module and MLP is preceded by a Layer Normalization (LN) layer, and followed by a residual connection. The LN layer is a normalization technique used to stabilize and accelerate the training process of neural networks. The calculation of two consecutive Swin Transformer blocks[19] is performed using:
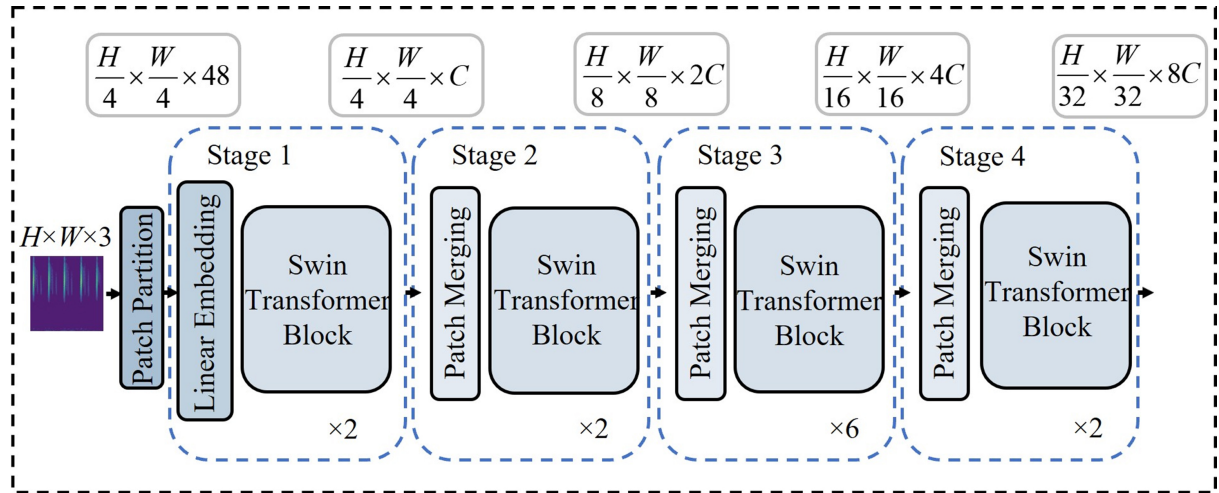
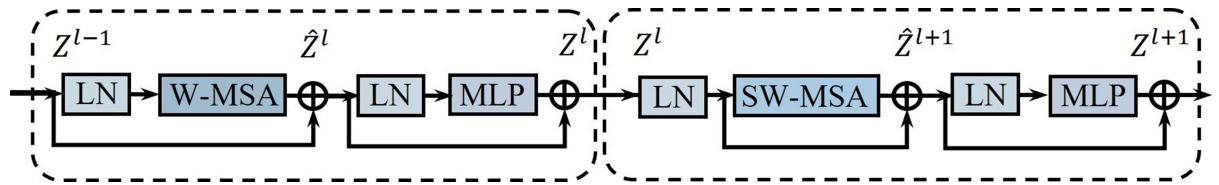**Figure 1.** Swin Transformer architecture diagram.

Figure 2. Diagram of Swin transformer block. LN: Layer Normalization; W-MSA: window-based MSA module, MSA: multi-head self-attention; MLP: Multilayer Perceptron; SW-MSA: shift window-based MSA module.

$$\hat{Z}^l = W\text{-}MSA(LN(Z^{l-1})) + Z^{l-1} \tag{1}$$

$$Z^l = MLP(LN(\hat{z}^l)) + \hat{Z}^l \tag{2}$$

$$\hat{Z}^{l+1} = SW\text{-}MSA(LN(Z^l)) + Z^l \tag{3}$$

$$Z^{l+1} = MLP(LN(\hat{Z}^{l+1})) + \hat{Z}^{l+1} \tag{4}$$

where $\hat{Z}^l$ and Z^l denote the output features of the (S)WMSA and MLP modules of block l, respectively.

Self-attention based on moving windows
Unlike traditional MSA, which performs complex calculations on the global image, W-MSA divides the image into non-overlapping windows. This method calculates each window separately, which significantly reduces computational complexity. However, the modeling capabilities of the system are limited due to the lack of cross-window connections. SW-MSA introduces cross-window connections, which increase the perceptual field of view through simple window shifting. This makes the effect more significant in image classification. Figure 3 demonstrates an efficient batch computation[19] using shift configuration to solve the multi-window problem caused by moving window partitions. To achieve this, the windows with different sizes can be combined to maintain the calculation amount. Then, the combined windows of the same size can be calculated and the calculation data to the original window can be transferred.
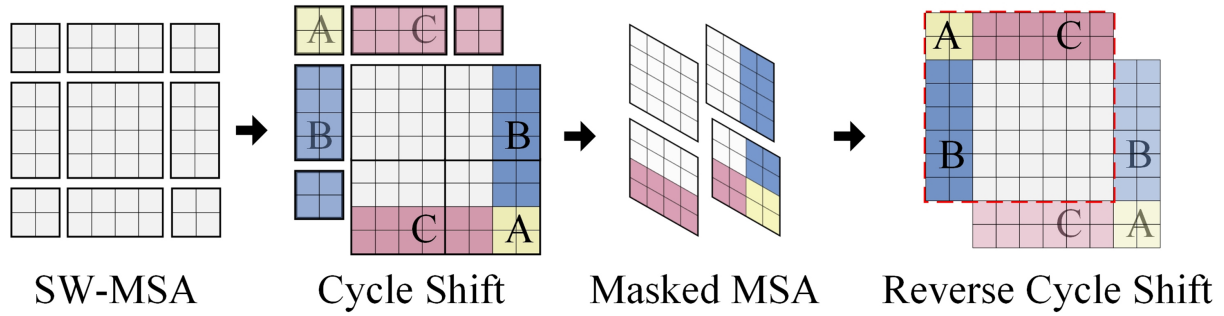
MSA[19] is performed using:

SW-MSA   Cycle Shift   Masked MSA   Reverse Cycle Shift

Figure 3. Example of SW-MSA calculation. MSA: Multi-head self-attention; SW-MSA: shift window-based MSA module.

$$Attention(Q,K,V) = SoftMax(QK^T / \sqrt{d} + B)V \tag{5}$$

where Q, K, $V \in R^{M^2 \times d}$ is the query matrix, key matrix, and value matrix; d is the query key dimension, and M^2 is the number of window patches; $B \in R^{M^2 \times M^2}$ is the relative position parameter, which is introduced similarly to position embedding in Transformer.

Continuous wavelet transform

The CWT[22] is a method for analyzing time-frequency information at multiple scales. The CWT decomposes the signal into a time-scale plane by scaling and shifting the base wave, transforming the one-dimensional vibration signal into a two-dimensional time-frequency diagram. This better represents the characteristics of the original signal. CWT is performed using[22]:

$$U(\alpha,\beta) = \int_{-\infty}^{+\infty} x(t)\overline{\psi}(t)dt = \frac{1}{\sqrt{|\alpha|}} \int_{-\infty}^{+\infty} x(t)\overline{\psi}(\frac{t-\beta}{\alpha})dt \tag{6}$$

where $U(\alpha, \beta)$ denotes the coefficients of the wavelet function, characterizing the similarity between the wavelet func Ation and the original signal; $\alpha, \beta \in R$ ($\alpha \neq 0$) are denoted as the scale parameter and translation parameter, respectively; $x(t)$ represents the original signal; $\overline{\psi}(t)$ indicates a wavelet basis function, and $\overline{\psi}(t)$ is the conjugate function of $\psi(t)$.

The wavelet basis function plays a pivotal role in the wavelet transform and is defined by five key properties: orthogonality (or bi-orthogonality), symmetry (or linear phase), regularity, vanishing moments, and tight support. Consequently, the wavelet basis functions are capable of performing multi-scale decomposition of signals, exhibiting satisfactory localization properties in both the time and frequency domains. This enables the wavelet transform to effectively capture transient changes and localized features in the signal. The choice of wavelet basis function is crucial for CWT. Commonly used wavelet basis functions include Haar and Morlet. In this article, the Morlet function is selected as the wavelet basis function. Its mathematical form[23] is performed using:

$$\psi(t) = e^{i\omega_0 t} e^{-\frac{t^2}{2}} \tag{7}$$

where ω_0 is the center frequency.

The proposed method

Figure 4 shows the process of the proposed damage identification method based on Swin Transformer. First, the initial vibration signals are collected from various damage conditions; next, sliding window is
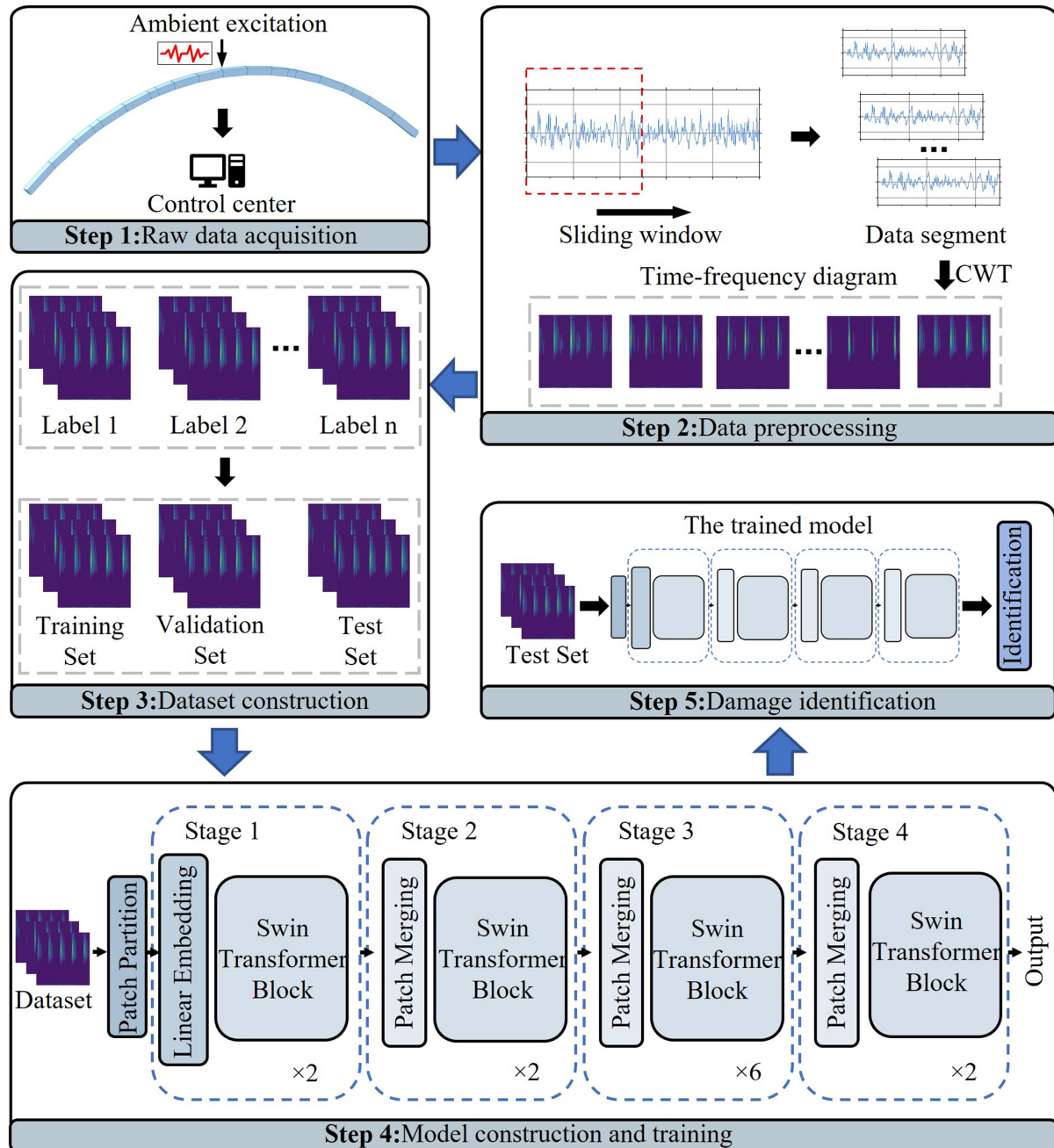
Figure 4. Flowchart of the method. CWT: Continuous wavelet transform.

applied to process the original data samples and divide them into smaller segments; then, the data is subsequently transformed into wavelet time-frequency diagrams through CWT. The resulting diagrams are labeled, shuffled, and divided into training, validation, and test sets in a 6:2:2 ratio; subsequently, the Swin Transformer model is trained using the training and validation sets. The model automatically extracts temporal and spatial features from the image during training; finally, the test set is inputted into the trained model to perform feature matching and complete damage identification.
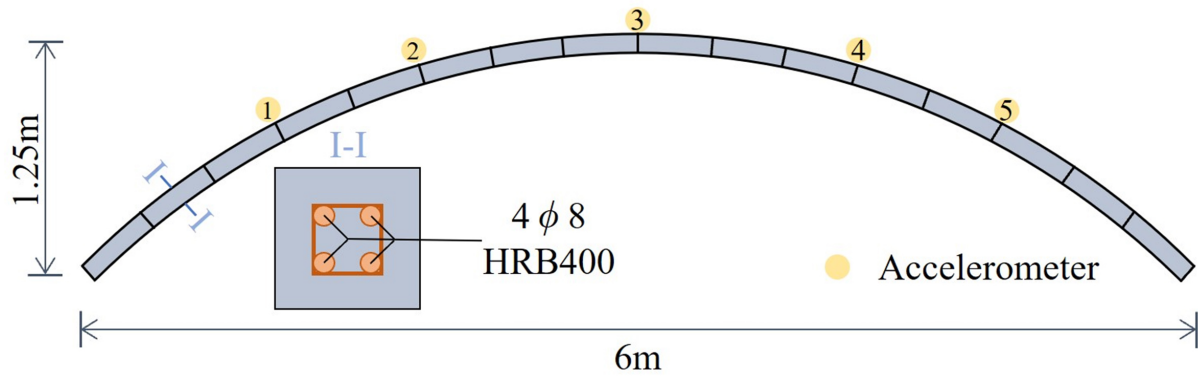
Figure 5. Finite element model.

NUMERICAL VALIDATION

Finite element model

To assess the effectiveness and feasibility of the proposed method, this paper establishes a numerical model to simulate various damage scenarios for identification. Figure 5 shows the model established by MIDAS finite element software. The arch structure is a single-rib rectangular cross-section arch made of C40 concrete. The span is 6 m, with a rise of 1.25 m. The arch axis coefficient is 1.9. The correctness of the model has been verified. To simulate structural damage, this model employs a stiffness reduction approach[24]. Specifically, the overall damage of the model is achieved by reducing the elastic modulus of concrete throughout the entire model by a certain percentage. Six damage scenarios with different degrees (Reduction ratio of elastic modulus) are presented in Table 1. The time-history load function is then applied as an excitation to the entire bridge by adding nodal dynamic loads near the mid-span of the numerical model. Thus, the response of the model in each state is obtained.

Dataset acquisition

In the numerical modeling, each damage scenario was sampled at a frequency of 256 Hz, and 240 s of acceleration time history data were collected, resulting in 61,440 data points [Figure 6]. Due to the need for a large amount of data in deep learning training, an overlapping sliding window[24] method is used to increase the number of samples [Figure 7]. Each window has a length of T = 512 data points, and the sliding step size is S = 120 data points, resulting in a total of N = 500 samples.

$$N = (L - T) / S + 1 \tag{8}$$

Where N is the number of samples obtained by cropping; L is the length of the original response; T is the length of the sliding window; S is the stride of the sliding window.

Subsequently, the one-dimensional time series data is transformed into a two-dimensional time-frequency diagram utilizing the Morlet wavelet in CWT, which depicts the distribution of signal energy at varying times and frequencies. Structural damage leads to alterations in the stiffness of the structure, subsequently affecting its natural frequencies. These frequency changes are manifested in the time-frequency diagram as the emergence or disappearance of specific frequency components. Furthermore, structural damage can result in a redistribution of vibration energy, evident as changes in the energy concentration areas within the time-frequency diagram. Additionally, structural damage may induce transient vibration characteristics, which appear as anomalies in specific localized time-frequency regions of the diagram. Consequently, the transformation of the local area of the time-frequency diagram is employed as a feature for the deep

Table 1. Damage scenarios

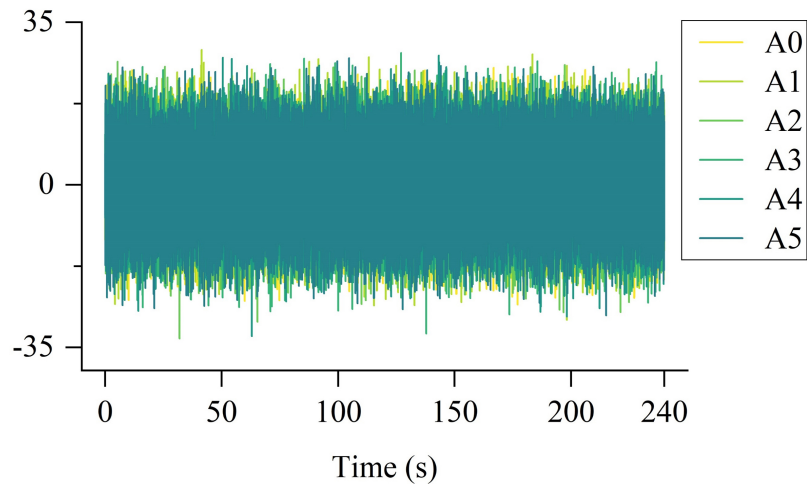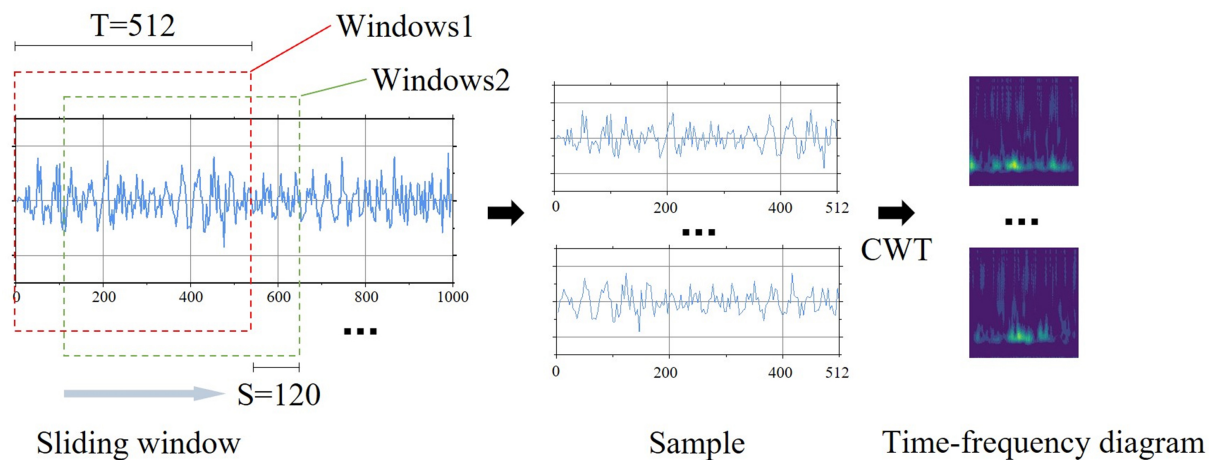| Scenarios	A0	A1	A2	A3	A4	A5
Damage degree	Undamaged	10%	20%	30%	40%	50%

Figure 6. Raw acceleration data.

**Figure 7.** Schematic diagram of data processing. CWT: Continuous wavelet transform.

learning model to learn about the damage. Figure 8 illustrates the time-frequency diagrams under different scenarios within the same time period. It can be observed that the changes induced by the damage are quite evident. After completing all conversions, the time-frequency diagrams are labeled with the damage scenarios and then divided into training, verification, and test sets according to a ratio of 6:2:2 [Table 2].

Model training and identification results

The Pytorch deep learning framework is used to compile the model. Model training and result identification are completed using a Windows 10 computer with an Intel Core i5-9300H CPU, GeForce GTX 1660-Ti GPU, and 16.00 GB of memory. The learning rate of the model is set to 0.005, batch size is 8, and epoch is 100, with weight decay of 1e-5 to prevent overfitting. The Swin Transformer model progressively extracts

Table 2. Details of damage scenarios and division of samples

Scenarios	Damage degree	Training sets	Validation set	Testing set	Total
A0	Undamaged	300	100	100	500
A1	10%	300	100	100	500
A2	20%	300	100	100	500
A3	30%	300	100	100	500
A4	40%	300	100	100	500
A5	50%	300	100	100	500

| A0 | A1 | A2 | A3 | A4 | A5 |
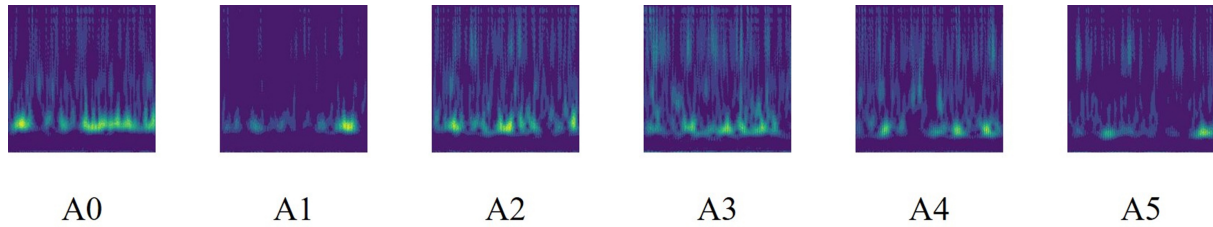
Figure 8. Time-frequency diagram of each scenario.

multi-scale features from images through a hierarchical structure. At each layer, the Swin Transformer partitions the input image using a sliding window mechanism and performs self-attention computations within each window. The self-attention mechanism allows the model to focus on important regions of the images, capturing changes in frequency and energy distribution in the time-frequency diagrams caused by structural damage, thereby completing the model training. After processing through multiple layers, the Swin Transformer aggregates the extracted features and performs classification through a fully connected layer, ultimately outputting the damage recognition results.

To demonstrate the superiority of the Swin Transformer used in this method, a comparison was made with InceptionV3, ResNet50, and CNN. Figure 9 shows the accuracy and loss curves of these deep learning models during the training process. From the figure, it can be observed that, except for CNN, the training performances of Swin Transformer, InceptionV3, and ResNet50 are quite good, with both accuracy and loss eventually converging. Although the accuracy and loss curves of the Swin Transformer on the training set do not fit as well as those of InceptionV3 and ResNet50, the curves on the validation set fit better than those of InceptionV3 and ResNet50. This indicates that the Swin Transformer does not overfit during the training process and has strong generalization capabilities. Therefore, it can be concluded that the Swin Transformer outperforms the other models in terms of performance. On the other hand, the accuracy and loss curves of CNN, both on the training and validation sets, do not converge, indicating that CNN is unsuitable for the dataset used in this method.

To evaluate the recognition effect of this model, accuracy, precision, recall, and F1 indicators are selected. The test results of the model are presented in Table 3.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

Table 3. The results of model test

Model	Accuracy (%)	Precision	Recall	F1
Swin Transformer	95.50	0.9553	0.9550	0.9549
InceptionV3	90.50	0.9116	0.9050	0.9054
ResNet50	89.00	0.9029	0.8900	0.8912
CNN	70.50	0.6927	0.7050	0.6925

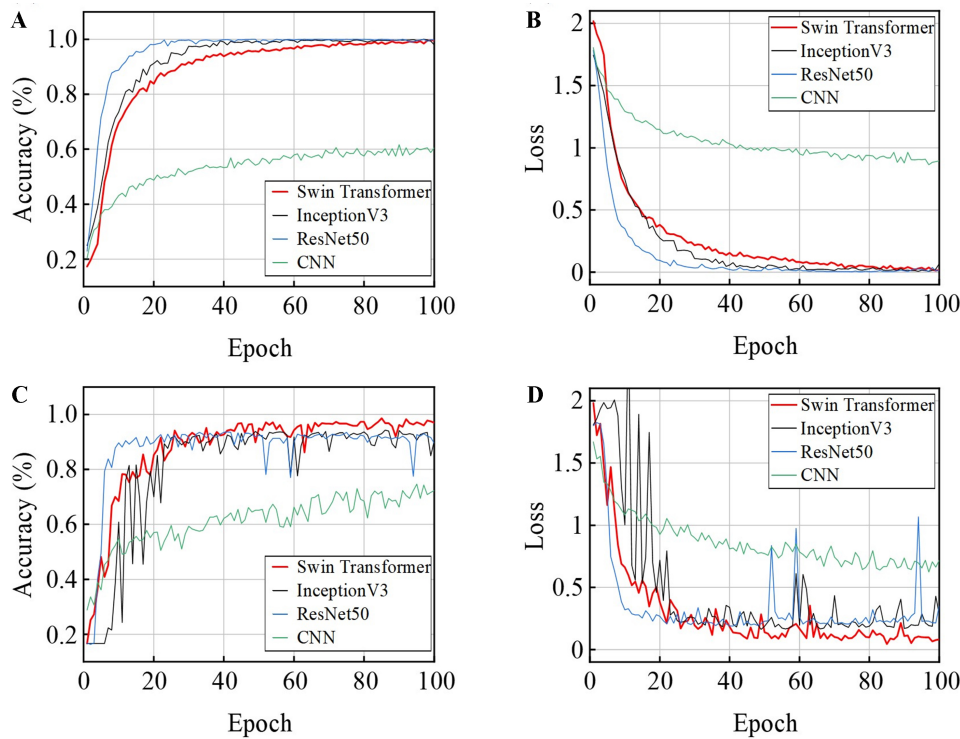CNN: Convolutional neural network.

Figure 9. Training results. (A) Accuracy of training set; (B) Loss of training set; (C) Accuracy of validation set; (D) Loss of validation set. CNN: Convolutional neural network.

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{12}$$

Where TP denotes true positive, i.e., the number of samples that are actually positive and are predicted to be positive; TN stands for true negative, which is the number of samples that are actually negative and are predicted to be negative; FP represents false positive, i.e., the number of samples that are actually negative but are predicted to be positive; FN points to false negative, which is the number of samples that are actually positive but predicted to be negative.

The test results in Table 3 further demonstrate the superiority of the Swin Transformer. With an accuracy of 95.5%, the Swin Transformer significantly outperforms InceptionV3, ResNet50, and CNN. Although InceptionV3 and ResNet50 also achieve decent recognition results, with accuracies around 90%, the Swin Transformer substantially improves the accuracy of damage identification. Additionally, the confusion
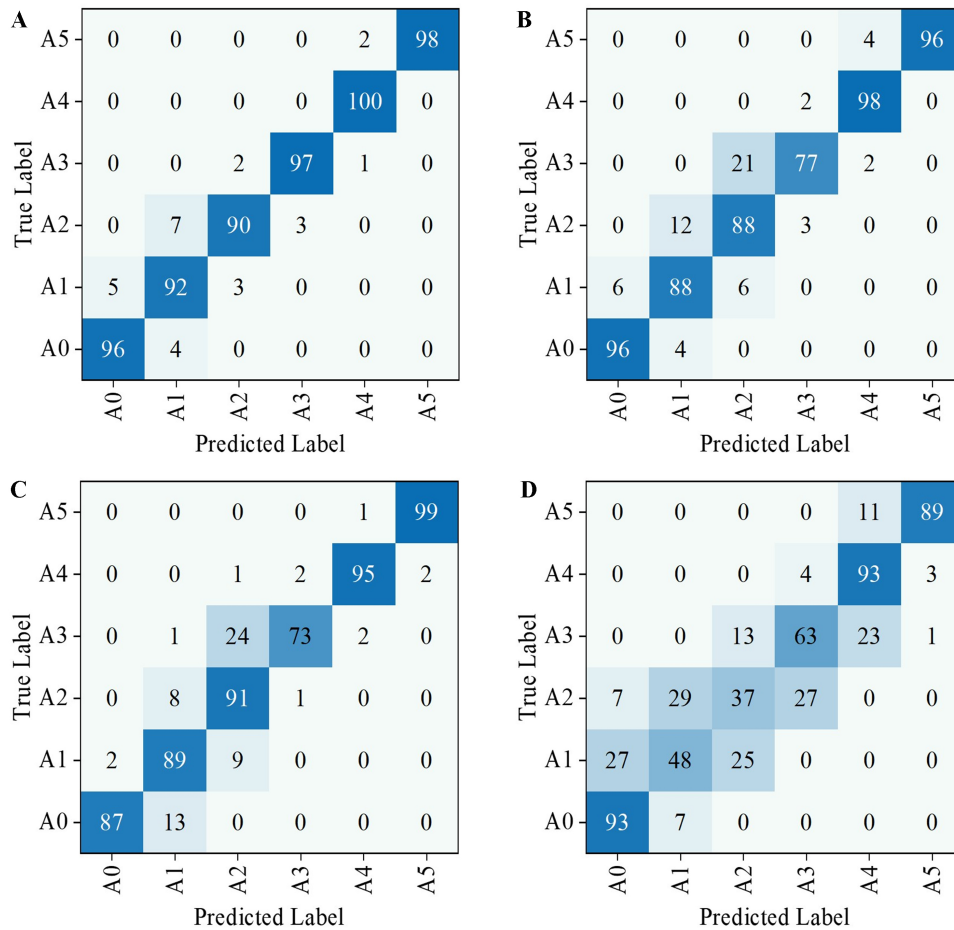
Figure 10. Confusion matrices. (A) Confusion matrix of Swin Transformer; (B) Confusion matrix of InceptionV3; (C) Confusion matrix of ResNet50;(D) Confusion matrix of CNN. CNN: Convolutional neural network.

matrix in Figure 10 shows that the Swin Transformer misclassifies only a small number of samples, whereas InceptionV3, ResNet50, and CNN misclassify relatively more samples, with CNN having the poorest recognition performance.

Robustness analysis
The feasibility and effectiveness of this method are demonstrated by the time history response data of the finite element model. Next, a robustness analysis was conducted to test the practical application potential of the model in terms of damage identification performance.

Sample length analysis
To investigate the effect of sampling frequency and sample length on damage identification, data is extracted for 240 and 120 s at sampling frequencies of 256 and 512 Hz, respectively. The data is divided into sample lengths of 256, 512, and 1,024.

Table 4 illustrates that the proposed method exhibits superior accuracy at both sample lengths of 512 and 1,024, while exhibiting slightly inferior performance at 256. As the length of the sample increases at a constant sampling frequency, the quantity of information contained in the data also increases. Consequently, the deep learning model can extract a greater number of features, thereby enhancing the

Table 4. The results of parameter analysis

Frequency	Sample length	Accuracy (%)	Precision	Recall	F1
256	256	88.67	0.8879	0.8867	0.8866
	512	95.50	0.9553	0.9550	0.9549
	1,024	98.83	0.9891	0.9883	0.9884
512	256	90.17	0.9042	0.9017	0.9023
	512	96.83	0.9689	0.9683	0.9682
	1,024	99.50	0.9951	0.9950	0.9950

accuracy of the recognition process. As the sampling frequency is increased, a greater quantity of information is captured simultaneously. Therefore, for a given sample length, a higher sampling frequency will result in a greater number of features being included in the collected data, and, consequently, an enhanced accuracy of recognition. In general, the proposed method demonstrated satisfactory performance in terms of accuracy across a range of sampling frequencies and sample lengths.

Noise analysis

There may be some noise in the actual operation environment. To simulate real-world conditions, a certain level of environmental noise is introduced. The addition of noise disturbs the actual distribution of samples, which can cause the model to identify the sample as the wrong type and reduce recognition accuracy. However, it can also test the ability of the model to withstand noise. Therefore, 10%, 20%, and 30% environmental noise are added to the acceleration time history response data with a sampling frequency of 256 Hz and a sample length of 512, and the noise in the actual measurement process is performed using[25]:

$$a_{noise} = a + RMS(a) \times N_{level} \times N_{unit} \tag{13}$$

where a_{noise} and a are the acceleration response data containing noise and the original data, respectively; $RMS(a)$ is the root mean square of a; N_{unit} is Gaussian white noise; and N_{level} is the added noise level, which is 10%, 20% and 30% in this paper.

From Table 5, it can be seen that the noise has little effect on the recognition accuracy, and the accuracy does not decrease significantly as the noise increases. This indicates that the proposed method has good noise immunity and can still achieve good recognition accuracy even under noise interference.

EXPERIMENTAL VALIDATION

Set up of the experiment

To verify the practicality of this method, this paper conducted a damage identification experiment on a reinforced concrete arch model [Figure 11]. This model was cast with C40 concrete; the arch rib section was rectangular, and the longitudinal bars were made of HRB400 rebar. The mass increase method was used to simulate damage. In this experiment, single damage (B0~B5) and multiple damage (C1~C6) were set up [Table 6]. Prefabricated steel plates were added to the load frame and their mass was increased using the lever principle. The loading device consisted of five I-beams installed at 1/6L, 1/3L, 1/2L, 2/3L and 5/6L. Due to the lever loading method, a tie rod was placed at the rear end of the loader to be fixed to the ground and the steel beam with bolts, and a limit rod fixed to the ground was placed to prevent horizontal deflection during loading. Meanwhile, a transverse limiter was installed in the center of the span to prevent the arch rib becoming unstable. The model was excited with a rubber hammer in the middle of the span after loading. At the same time, the vertical accelerometer was used to collect acceleration response data under various damage for 240 s at a sampling rate of 256 Hz.

Table 5. The results of noise analysis

Noise level	Accuracy (%)	Precision	Recall	F1
No noise	95.50	0.9553	0.9550	0.9549
10%	95.17	0.9520	0.9517	0.9518
20%	95.17	0.9521	0.9517	0.9517
30%	94.83	0.9489	0.9483	0.9480

Table 6. Damage scenarios

Scenarios	Location of damage
B0	Undamaged
B1	1/6L
B2	1/3L
B3	1/2L
B4	2/3L
B5	5/6L
C1	1/6L, 1/3L
C2	1/6L, 1/2L
C3	1/6L, 2/3L
C4	1/6L, 5/6L
C5	1/3L, 1/2L
C6	1/3L, 2/3L

Figure 11. Experimental model. (A) Unloaded model; (B) 1/3L, 1/2L loading damage; (C) Accelerometers and arch loading point.

Damage identification

For the acceleration response data collected under each scenario, 500 time-frequency diagrams were obtained, where the sliding window with a length of 512 and stride of 120 is selected for CWT[24]. Subsequently, the data is labeled and disordered to construct the training, validation and test sets in the ratio of 6:2:2. The training and validation sets are then fed into the model for training, and the test set is fed into the model for damage recognition when training is complete. The training results and confusion matrices are shown in Figures 12 and 13, and the recognition accuracy is shown in Table 7. From the table, it can be seen that the proposed method is very effective in damage identification both in single damage and multi-damage, with the accuracy of 99.6% and 99.0%, respectively, which further illustrates the practicability of this method.

Table 7. Damage identification results

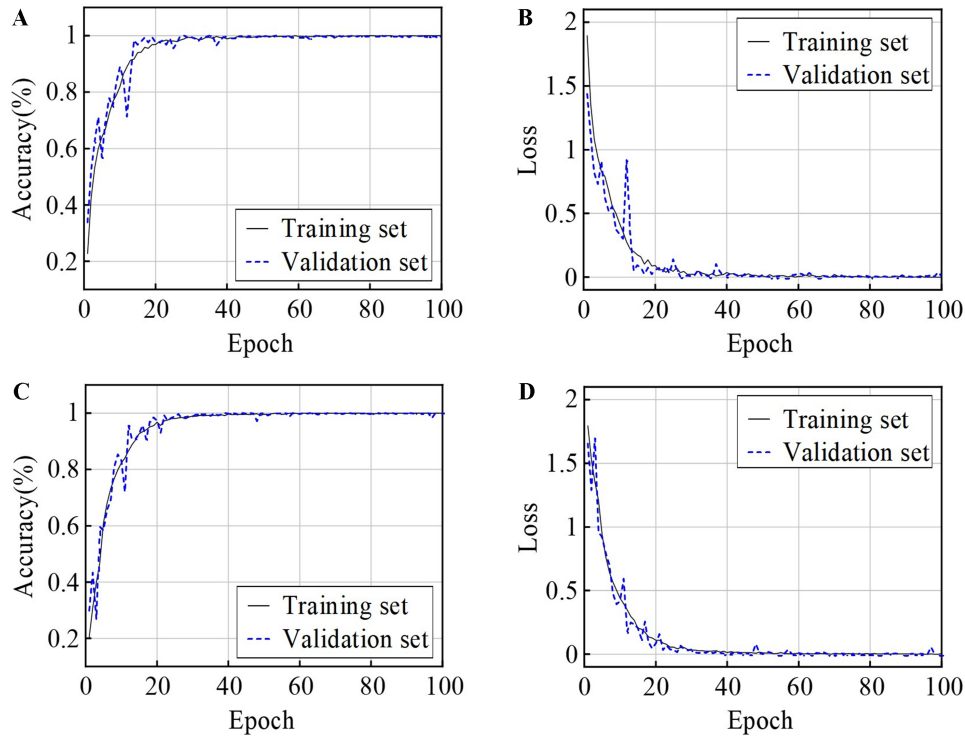| Scenarios	Accuracy (%)	Precision	Recall	F1
Single damage	99.6	0.9966	0.9966	0.9966
Multi damage	99.0	0.9902	0.9900	0.9899

Figure 12. Training results of damage identification. (A) Accuracy of single-damage training and validation sets; (B) Loss of single-damage training and validation sets; (C) Accuracy of multi-damage training and validation sets; (D) Loss of multi-damage training and validation sets.
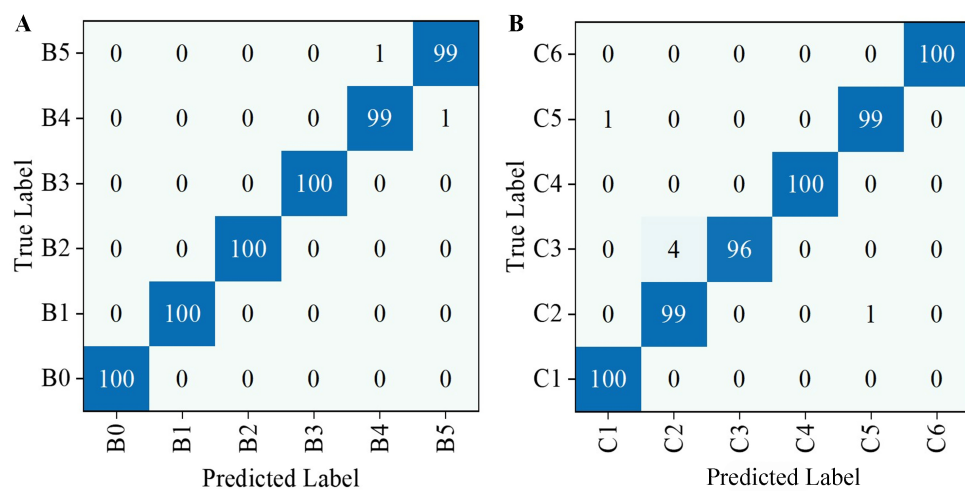

Figure 13. Confusion matrices of the test result. (A) Single- damage confusion matrix; (B) Multi- damage confusion matrix.

CONCLUSIONS

In this paper, a novel structural damage identification method based on Swin Transformer and CWT is proposed. The effectiveness and practicality of the proposed method are verified numerically and experimentally. Some main conclusions can be obtained:

(1) The combination of Swin Transformer and CWT can be used as an effective approach for damage identification. This method can accurately identify different damage scenarios, obtaining an identification accuracy of 96% in numerical simulation.

(2) The proposed method presents high robustness. In the situations with the disturbance of noise, the test accuracy of all scenarios exceeds 95%. The recognition accuracy under different sampling frequencies and sample lengths is more than 94%.

(3) The proposed method has high practicability. The experimental test has yielded excellent recognition performance, with recognition accuracy surpassing 99% for both single and multiple damage scenarios. Therefore, this method may have a noticeable potential in practice.

DECLARATIONS

Authors' contributions
Made substantial contributions to conception and design of the study and conducted data analysis and interpretation: Xin J, Tao G, Tang Q
Performed data acquisition and provided administrative, technical, and material support: Zou F, Xiang C

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

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

REFERENCES

1. He Z, Li W, Salehi H, Zhang H, Zhou H, Jiao P. Integrated structural health monitoring in bridge engineering. *Automat Constr* 2022;136:104168. DOI

2. Khodabandehlou H, Pekcan G, Fadali MS. Vibration-based structural condition assessment using convolution neural networks. *Struct Control Health Monit* 2019;26:e2308. DOI

3. Gonen S, Erduran E. A hybrid method for vibration-based bridge damage detection. *Remote Sensing* 2022;14:6054. DOI

4. Daneshvar MH, Saffarian M, Jahangir H, Sarmadi H. Damage identification of structural systems by modal strain energy and an optimization-based iterative regularization method. *Eng Comput* 2023;39:2067-87. DOI

5. Pooya SMH, Massumi A. A novel damage detection method in beam-like structures based on the relation between modal kinetic energy and modal strain energy and using only damaged structure data. *J Sound Vib* 2022;530:116943. DOI

6. An X, Zhang Q, Li C, Hou J, Shi Y. Damage identification of semi-rigid joints in frame structures based on additional virtual mass method. *Sensors* 2022;22:6495. DOI PubMed PMC

7. Zhang Y, Xiong Z, Liang Z, She J, Ma C. Structural damage identification system suitable for old arch bridge in rural regions: random forest approach. *Comp Model Eng Sci* 2023;136:447-69. DOI

8. Gomes GF, de Almeida FA, Junqueira DM, da Cunha SS, Ancelotti AC. Optimized damage identification in CFRP plates by reduced mode shapes and GA-ANN methods. *Eng Struct* 2019;181:111-23. DOI

9. Cuong-Le T, Nghia-Nguyen T, Khatir S, Trong-Nguyen P, Mirjalili S, Nguyen KD. An efficient approach for damage identification based on improved machine learning using PSO-SVM. *Eng Comput* 2022;38:3069-84. DOI

10. Ren J, Zhang B, Zhu X, Li S. Damaged cable identification in cable-stayed bridge from bridge deck strain measurements using support vector machine. *Adv Struct Eng* 2022;25:754-71. DOI

11. Farias SV, Saotome O, Campos Velho HF, Shiguemori EH. A damage detection method using neural network optimized by multiple particle collision algorithm. *J Sensors* 2021;2021:1-14. DOI

12. Tang Q, Zhou J, Xin J, Zhao S, Zhou Y. Autoregressive model-based structural damage identification and localization using convolutional neural networks. *KSCE J Civ Eng* 2020;24:2173-85. DOI

13. Mai HT, Lee S, Kang J, Lee J. A damage-informed neural network framework for structural damage identification. *Comput Struct* 2024;292:107232. DOI

14. Mousavi M, Gandomi AH. Structural health monitoring under environmental and operational variations using MCD prediction error. *J Sound Vib* 2021;512:116370. DOI

15. Sony S, Gamage S, Sadhu A, Samarabandu J. Vibration-based multiclass damage detection and localization using long short-term memory networks. *Structures* 2022;35:436-51. DOI

16. Fu L, Tang Q, Gao P, Xin J, Zhou J. Damage identification of long-span bridges using the hybrid of convolutional neural network and long short-term memory network. *Algorithms* 2021;14:180. DOI

17. Fernandez-Navamuel A, Pardo D, Magalhães F, Zamora-Sánchez D, Omella ÁJ, Garcia-Sanchez D. Bridge damage identification under varying environmental and operational conditions combining Deep Learning and numerical simulations. *Mech Syst Signal Proc* 2023;200:110471. DOI

18. Sony S, Gamage S, Sadhu A, Samarabandu J. Multiclass damage identification in a full-scale bridge using optimally tuned one-dimensional convolutional neural network. *J Comput Civ Eng* 2022;36:04021035. DOI

19. Liu Z, Lin Y, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 10-17; Montreal, Canada. IEEE; 2021. pp. 9992-10002. DOI

20. Üzen H, Türkoğlu M, Yanikoglu B, Hanbay D. Swin-MFINet: Swin transformer based multi-feature integration network for detection of pixel-level surface defects. *Expert Syst Appl* 2022;209:118269. DOI

21. Xu Y, Wang X, Zhang H, Lin H. SE-Swin: an improved Swin-Transfomer network of self-ensemble feature extraction framework for image retrieval. *IET Image Process* 2024;18:13-21. DOI

22. Miao R, Shan Z, Zhou Q, et al. Real-time defect identification of narrow overlap welds and application based on convolutional neural networks. *J Manuf Syst* 2022;62:800-10. DOI

23. Zhou J, Li Z, Chen J. Application of two dimensional Morlet wavelet transform in damage detection for composite laminates. *Compos Struct* 2023;318:117091. DOI

24. Hou Y, Qian S, Li X, Wei S, Zheng X, Zhou S. Application of vibration data mining and deep neural networks in bridge damage identification. *Electronics* 2023;12:3613. DOI

25. Diao Y, Men X, Sun Z, Guo K, Wang Y. Structural damage identification based on the transmissibility function and support vector machine. *Shock Vib* 2018;2018:1-13. DOI