Transport and Communications Science Journal, Vol. 77, Issue 4 (05/2026), 485-499

Transport and Communications Science Journal

# VIBRATION-BASED DAMAGE DETECTION IN CABLE-STAYED BRIDGES USING A NOVEL 1D-CONVNEXT-LSTM NETWORK

**Ho Xuan Nam, Vu Manh Trung, Nguyen Nam Son, Nguyen Ngoc Long***

University of Transport and Communications, No 3 Cau Giay Street, Hanoi, Vietnam

ARTICLE INFO

TYPE: Research Article
Received: 26/03/2026
Revised: 23/04/2026
Accepted: 27/04/2026
Published online: 15/05/2026
https://doi.org/10.47869/tcsj.77.4.11

* *Corresponding author*
Email: nguyenngoclong@utc.edu.vn; Tel: 0913381128

**Abstract.** Structural health monitoring (SHM) plays a crucial role in maintaining the safety and serviceability of civil infrastructure, such as cable-stayed bridges. However, simultaneously extracting detailed local signal features and long-range temporal dependencies from vibration data remains a significant challenge for standalone deep learning architectures like LSTM or ResNet1D. To address this limitation, this study proposes an advanced hybrid architecture, 1D-ConvNeXt-LSTM, for structural damage detection. This framework integrates the multi-scale feature extraction capabilities of 1D-ConvNeXt with the sequential modeling proficiency of Long Short-Term Memory (LSTM) networks. The proposed method was evaluated using vibration time-series data acquired from a scaled cable-stayed bridge model under five distinct simulated damage scenarios. Experimental results demonstrate that the 1D-ConvNeXt-LSTM model achieves superior classification performance, yielding a macro–Area Under the Curve (AUC) of 0.971 and a macro F1-score of 0.814, significantly outperforming baseline architectures including FCN, ResNet1D, and LSTM. Ultimately, the proposed architecture provides a robust, stable, and highly accurate solution for structural condition assessment, thereby enhancing the effectiveness of damage identification in practical SHM applications.

**Keywords:** cable-stayed bridge; time-series data; structural health monitoring; damage detection; one dimensional-convnext; long short-term memory.

## 1. INTRODUCTION

Structural health monitoring (SHM) plays a crucial role in maintaining the safety and serviceability of civil infrastructure such as bridges, buildings, and transportation systems [1]. SHM systems enable continuous observation of how structures behave during use by combining sensing technologies, data acquisition platforms, and analytical tools. The collected information allows engineers to evaluate structural conditions, identify abnormal responses, and support maintenance decision-making throughout the service life of the structure. In bridge engineering, long-term monitoring has become particularly important because bridges are subjected to complex traffic loads and environmental influences. Continuous monitoring data therefore provide an essential basis for assessing structural performance and improving infrastructure management strategies [2].

Current SHM practices utilize various sensing devices at important locations within a structure, which may consist of common types of monitoring devices such as accelerometers, strain sensors, displacement sensors, temperature sensors, and vibration sensors [3], [4]. These devices monitor the responses of a structure to various dynamic excitations. Consequently, SHM practices produce massive amounts of data that illustrate time-series data for understanding changes in responses over time. Several SHM practices, such as modal analysis, vibration monitoring, damage detection, and structural state assessment, are enabled by such data. However, it is important to note that the accuracy of the collected data is a fundamental requirement for the effectiveness of various SHM practices.

In vibration-based SHM systems, large volumes of time-series data are continuously collected from distributed sensors installed on structures. These measurements contain valuable information about the dynamic behaviour of the structure and can be used to identify potential structural damage. However, the interpretation of such data remains challenging due to the complex and nonlinear characteristics of structural responses as well as the influence of environmental and operational variations. These factors often obscure damage-related patterns in monitoring data and make reliable damage detection difficult [5], [6].

Traditional vibration-based damage detection methods commonly rely on signal processing techniques or physics-based indicators derived from modal parameters, frequency shifts, or statistical features [7]. Examples include modal analysis, frequency-domain indicators, and various statistical approaches designed to identify abnormal structural responses [8], [9], [10]. While these methods have demonstrated effectiveness in controlled environments, their performance may degrade when dealing with large-scale monitoring data and complex structural behaviour.

With the rapid advancement of artificial intelligence, machine learning (ML) and deep learning (DL) techniques have increasingly been applied to structural damage detection problems. These approaches are capable of automatically learning complex patterns directly from raw monitoring data without requiring explicit feature engineering. Various ML algorithms, such as support vector machines, k-nearest neighbours, and artificial neural networks, have been investigated for identifying structural damage patterns from vibration signals [11]. More recently, deep learning models have demonstrated strong potential in capturing nonlinear relationships and extracting hierarchical features from large-scale structural monitoring datasets.

Among these approaches, recurrent neural networks, particularly Long Short-Term Memory (LSTM) networks, have been widely applied for reconstructing SHM data because

they can effectively learn temporal dependencies in sequential signals [12]. However, LSTM-based models mainly focus on temporal relationships and may have limited capability in extracting detailed local features directly from raw signals. Convolution-based models such as ResNet1D have therefore been introduced to improve feature extraction through deep residual learning. Although ResNet1D can effectively learn hierarchical representations, its primary strength lies in capturing local signal patterns, and it may not fully model long-range temporal dependencies in sequential monitoring data [13].

To address these limitations, hybrid architectures that combine convolutional feature extraction with temporal modelling have been increasingly explored. The recently developed 1D-ConvNeXt architecture incorporates modern convolutional design principles that improve representation capability and training stability compared with traditional CNN or ResNet-based models [14]. By integrating 1D-ConvNeXt with LSTM, a complementary framework can be established in which the convolutional component captures local signal patterns and multi-scale features, while the LSTM layer models long-term temporal dependencies across the extracted feature sequences. Compared with standalone LSTM or ResNet1D models, the combined 1D-ConvNeXt–LSTM architecture provides a more balanced representation of both local signal characteristics and temporal dynamics, which can enhance the performance of time-series analysis and damage detection in SHM applications.

The remainder of this paper is organized as follows. Section 2 describes the proposed 1D-ConvNeXt–LSTM framework, while Section 3 introduces the dataset and experimental setup. Section 4 presents and discusses the experimental results, and Section 5 concludes the paper.

## 2. METHODOLOGY

### 2.1. Long Short-Term Memory Networks

LSTM networks are a specialized form of recurrent neural networks (RNNs) designed to effectively model sequential data and long-term temporal dependencies. Traditional RNNs often encounter the vanishing or exploding gradient problem when learning from long sequences, which limits their capability to preserve information over extended time intervals. LSTM networks address this issue by introducing a memory cell structure and a gating mechanism that regulates the flow of information through the network. This design allows the model to selectively retain or discard information across time steps, thereby improving the ability to learn long-range temporal relationships in sequential data [15].

An LSTM cell (Figure 1) consists of three main gates: the input gate, forget gate, and output gate. The forget gate determines which information from the previous cell state should be discarded, while the input gate controls the amount of new information that is written into the cell state. The output gate then regulates how much of the stored information is propagated to the hidden state at the current time step. Through these gated operations, LSTM networks maintain relevant temporal information while filtering out irrelevant signals, making them highly suitable for time-series analysis [16].

The key gating operations of LSTM can be expressed as follows:

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

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (2)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (3)$$

$$\tilde{c}_t = \tanh\left(W_c[h_{t-1}, x_t] + b_c\right) \quad (4)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (5)$$

$$h_t = o_t \odot \tanh\left(c_t\right) \quad (6)$$

where $f_t$, $i_t$, and $o_t$ represent the forget gate, input gate, and output gate, respectively. $h_t$ and $c_t$ denote the hidden state and memory cell state. These mechanisms allow LSTM to capture long-term temporal relationships in vibration signals generated by structural responses.

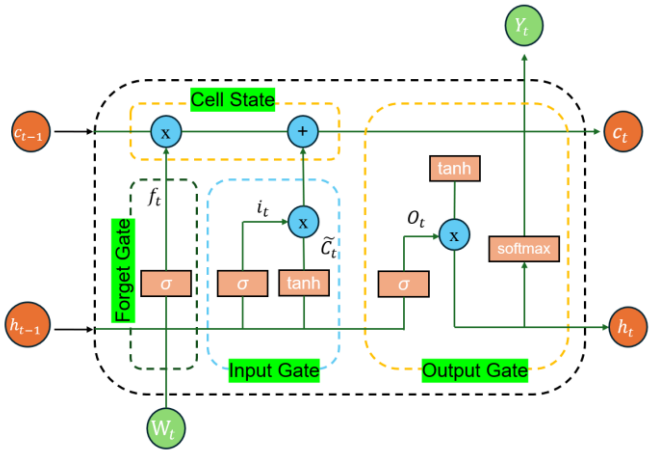



Figure 1: A structural diagram of an LSTM Cell.

## 2.2. One-Dimensional ConvNeXt networks

ConvNeXt is a modern convolutional neural network architecture designed to revisit and improve traditional convolutional networks by incorporating several design principles inspired by transformer-based models while maintaining the efficiency of convolution operations. The architecture introduces large-kernel depthwise convolutions, simplified residual blocks, and layer normalization, which collectively enhance the representation capability of convolutional networks and enable them to achieve competitive performance in various deep learning tasks [17].

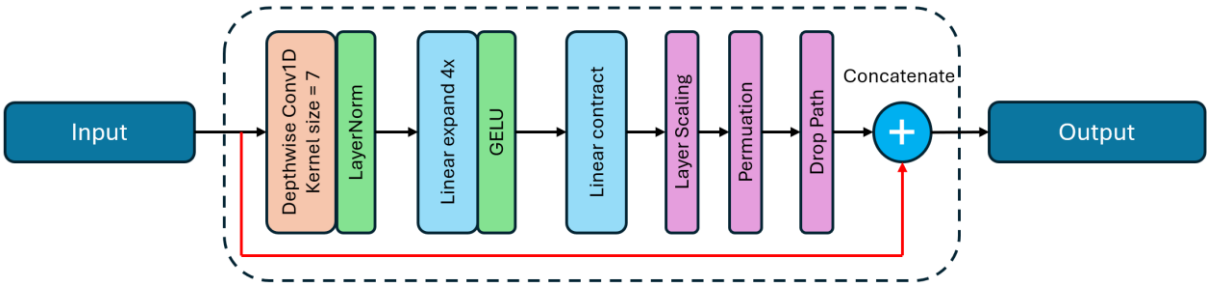



Figure 2: Structural diagram of a ConvNeXt1D block.

To enable the architecture to process sequential signals, ConvNeXt can be adapted from two-dimensional image processing to a one-dimensional convolutional structure, commonly referred to as ConvNeXt1D (Figure 2). In this configuration, the original two-dimensional convolution kernels are replaced with one-dimensional kernels that operate along the temporal dimension of the input signal. This modification allows the network to learn hierarchical

representations from time-series data while preserving the key design characteristics of ConvNeXt, including depthwise convolution, layer normalization, and residual connections:

$$y_c(t) = \sum_{k=0}^{K-1} w_{c,k}\, x_c(t + k - p) \quad (7)$$

where $x_c\,(t)$ represents the input signal of channel $c$ at time $t$, $w_{c,k}$ denotes the convolution kernel weights, $K$ is the kernel size, and $p$ represents the padding.

### 2.3. Proposed methodology for damage detection

To address the limitations of standalone convolutional networks in capturing long-range dependencies and recurrent networks in hierarchical feature learning, this study introduces the 1D-ConvNeXt-LSTM. This hybrid deep learning architecture is specifically designed for vibration-based structural damage detection and consists of five integrated components: an input stem, ConvNeXt stages, LSTM layers, dual-pooling aggregation, and a classification head.

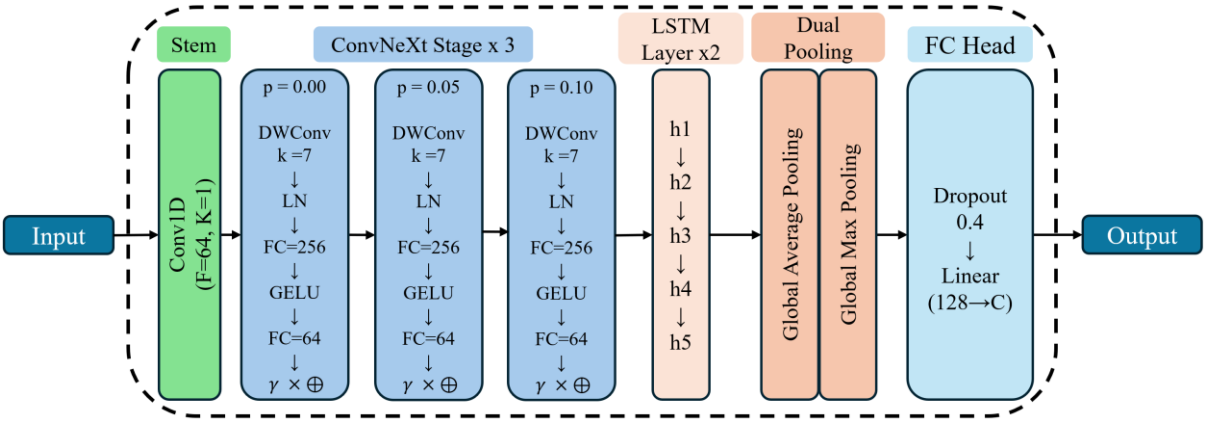



Figure 3: Architecture of proposed model.

The architecture begins with an input stem that maps raw multi-sensor vibration signals into a higher-dimensional feature space. These signals are subsequently processed through ConvNeXt1D stages, which consist of stacked blocks incorporating large-kernel depthwise convolutions, layer normalization, GELU activations, and residual connections. To maximize the effective receptive field, each stage is specifically configured with a kernel size of k = 7. Our preliminary sensitivity analysis indicated that this k = 7 setup is more effective than standard smaller kernels in capturing the intricate spatial-temporal characteristics required to detect subtle, damage-induced anomalies within the complex vibration time-series data.

This configuration efficiently extracts complex, hierarchical local patterns and structural response characteristics from the raw data. The high-level feature sequences are subsequently fed into stacked LSTM layers. By leveraging gated memory mechanisms, the LSTM modules capture long-term temporal dynamics and sequential dependencies while effectively filtering out monitoring noise.

Finally, a dual-pooling strategy fuses global average pooling (GAP) and global max pooling (GMP) to aggregate complementary features. This comprehensive representation is passed through a fully connected classification head utilizing dropout regularization to output the robust, final structural damage prediction

## 3. DATASET AND PREPROCESSING

### 3.1. Introduction to the cable-stayed bridge data acquisition model

This study utilizes a scaled cable-stayed bridge model installed in Laboratory A3 (Figure 4) at the University of Transport and Communications. The model has a total length of 3.76 m and consists of a steel girder with a cross-section of 0.12 m × 0.005 m and an A-shaped steel tower with a height of 1.6 meter. The structural system of the bridge is configured with a single vertical cable plane aligned along the girder centerline, comprising six stay cables symmetrically anchored in a fan arrangement. The support system includes spring supports at both ends of the girder and an elastic rubber bearing located at the tower's cross-beam position.

Figure 4: A3 cable-stayed bridge model.

To ensure the stability of the boundary conditions and to minimize vibration disturbances during dynamic measurements, the bridge tower and the girder-end supports were rigidly anchored to the concrete foundation (M350) using a system of high-strength bolts. In addition, mass attachment devices were integrated at the cable anchorage points on the girder, allowing flexible adjustment of system parameters for different experimental scenarios.

a) Vibration measurement sensor PCB 353B34.

b) Data acquisition unit NI cDAQ-9178 + Module NI-9234.

c) Signal cable PCB – Model: 003C20.

d) Dedicated vibration monitoring workstation.

Figure 5. Bridge vibration measurement equipment.

To collect vibration data from the model, high-sensitivity differential piezoelectric accelerometers (Model 353B34, PCB Piezotronics) operating under the ICP® standard were

employed (Figure 5a). A set of eight sensors was used, each with a nominal sensitivity of approximately 10 mV/(m/s²), a measurement range of ±491 m/s², and a linear frequency response range from 1 to 4000 Hz (±5%). The vibration signals from the sensors were transmitted through specialized low-noise cables (Model 003C20) (Figure 5c) to a central data acquisition system from National Instruments (NI). The DAQ system consists of an NI cDAQ-9178 (Figure 5b) chassis integrated with NI-9234 signal acquisition modules. The entire process of monitoring, recording, and real-time data processing was performed on a dedicated computer using the NI LabVIEW software platform (Figure 5d).

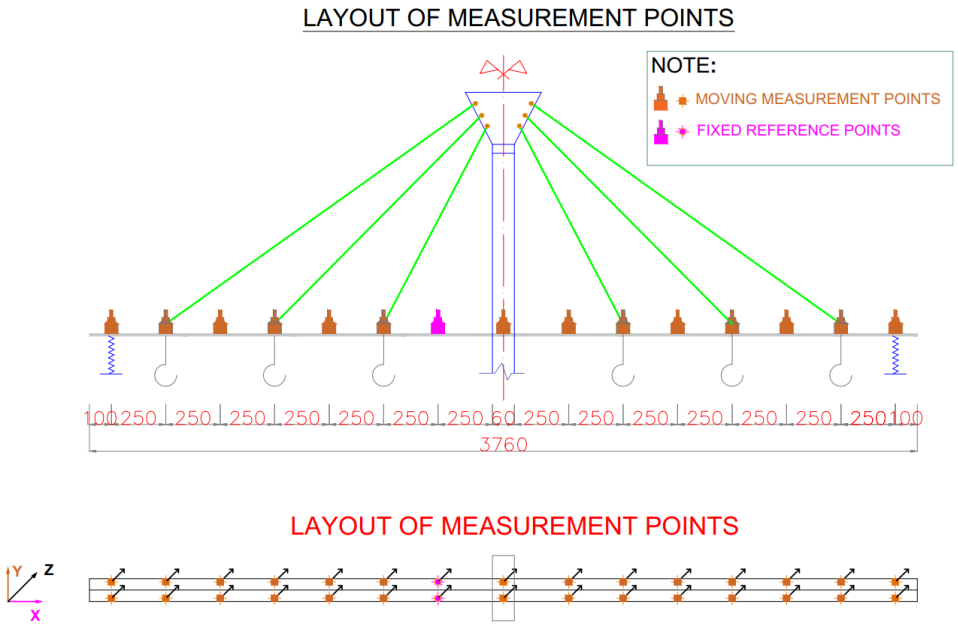



Figure 6: Layout of measurement points.

Due to the limited number of sensors (eight in total), this study conducted measurements using five different setups to collect vibration data for the entire bridge, with two reference points.

Table 1. Mass values (kg) corresponding to the five loading positions in each loading scenario.

| **Label** | **Location** | | | | | | **State** |
|---|---|---|---|---|---|---|---|
| | **01** | **02** | **03** | **04** | **05** | **06** | |
| **00** | 00 | 00 | 00 | 00 | 00 | 00 | Intact |
| **01** | 20 | 10 | 10 | 10 | 10 | 10 | Damaged |
| **02** | 10 | 10 | 30 | 10 | 10 | 10 | |
| **03** | 20 | 10 | 20 | 10 | 20 | 10 | |
| **04** | 20 | 10 | 20 | 10 | 10 | 10 | |

These setups were employed to record vibration responses under five different structural damage scenarios. The damage conditions were simulated by applying concentrated loads at locations considered representative of structural damage. Physically, attaching mass alters the system's dynamic properties by modifying the mass matrix $[M]$. From a structural dynamics

perspective, the natural frequency of a system is governed by the relationship $f = \frac{1}{2\pi}\sqrt{\frac{[K]}{[M]}}$. According to the fundamental vibration equation, an increase in mass leads to a reduction in natural frequencies, which mimics the frequency degradation typically observed when structural stiffness $[K]$, decreases due to cracks or cable tension loss. Although the underlying physical mechanisms differ, the resulting alterations in the vibration signatures - specifically frequency shifts and mode shape variations - are comparable. The damage cases are summarized in Table 1, with an example configuration illustrated in Figure 7. The measurement duration for each setup was 10 minutes.

Figure 7. Sample of damage cases in the cable-stayed bridge model (Label 01).

To obtain the natural vibration frequencies of the bridge, the experiment uses a rubber hammer striking at one point as the primary excitation source. The model has a total of 38 measurement points. Each sensor records 400,000 data points, and measurements are conducted under five damage scenarios. Therefore, the original dataset has the shape (5; 38; 400,000) illustrated in Figure 8.

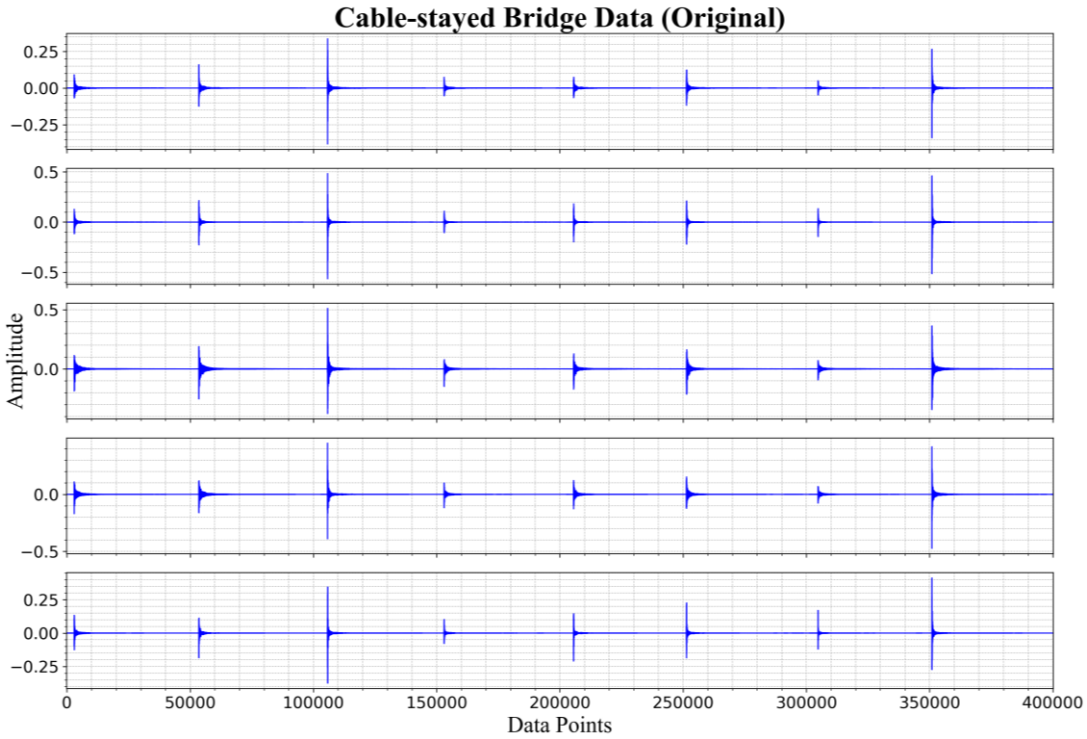



Figure 8. Sample of original input data.

### 3.2. Data preprocessing

To enhance data diversity and improve model generalization, several data augmentation (AUG) techniques were applied, including Gaussian noise injection, amplitude scaling, and signal reversal. For each original structural scenario, 5 augmented samples were generated using the augmentation techniques described above, producing an AUG dataset with a dimension of (25; 38; 400,000).

Because the vibration signals are very long, the AUG dataset was further partitioned into smaller sequences to facilitate efficient training and to evaluate the capability of the 1D-ConvNeXt–LSTM architecture in processing long time-series data and capturing temporal dependencies in structural vibration signals. Each sensor signal was divided into segments of 4,000 time steps, producing 100 segments per sensor. The total number of samples becomes 30 (scenarios) × 38 (samples per scenario) = 1,140 samples. Therefore, the final segmented dataset (Figure 9) has a dimension of (1140; 100; 4,000), where 100 represents the number of segments in each sample and 4,000 denotes the length of each segment. Finally, the dataset is divided into two sets: training (80%) and valid (20%).

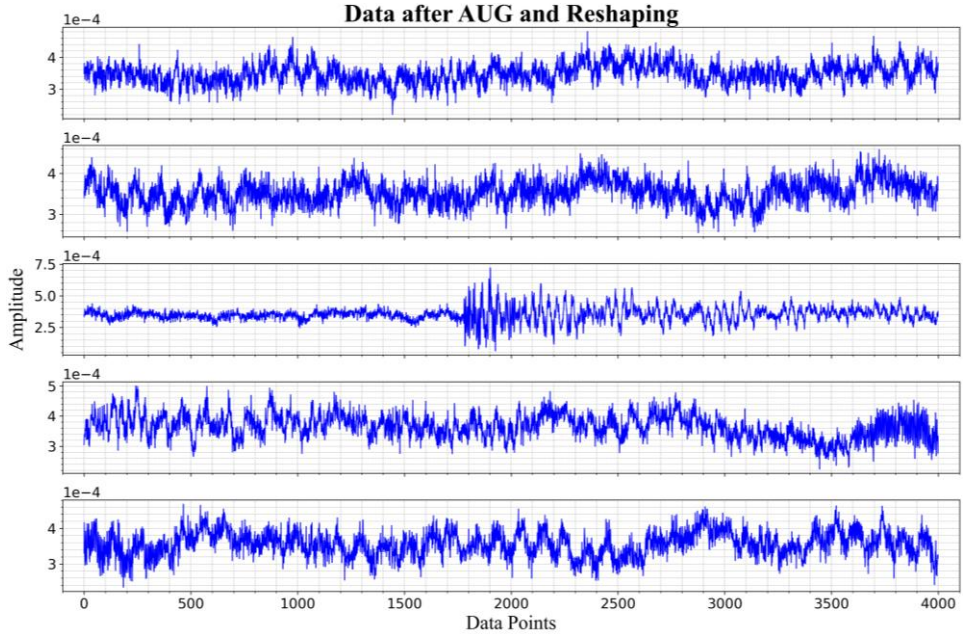



Figure 9. Augmented and reshaped dataset used for model training.

## 4. EXPERIMENTAL RESULTS

To benchmark the proposed 1D-ConvNeXt-LSTM against established paradigms, we compared it with FCN and ResNet1D (representing purely convolutional architectures for local feature extraction) and LSTM (representing recurrent networks for long-term temporal dependencies). All experiments were executed using the PyTorch framework on a workstation equipped with an Intel Core i9-13900HX CPU, 16 GB of RAM, and an NVIDIA RTX 4060 Laptop GPU.

To ensure equitable comparisons, all models were trained for 100 epochs with a batch size of 32 using the Adam optimizer, an initial learning rate of 0.001, and a weight decay of $1 \times 10^{-3}$. The combination of Focal Loss and Label Smoothing was strategically chosen to enhance the model's robustness. While the dataset is balanced across scenarios, certain

structural conditions exhibit high similarity in vibration patterns, creating hard-to-classify boundary cases. Focal Loss ($\gamma = 2$) was employed to prioritize learning from these difficult samples. Concurrently, a label smoothing factor of 0.15 was applied to prevent over-fitting and improve generalization by smoothing the target distributions.

All computational experiments were executed on a CUDA-enabled GPU utilizing the PyTorch deep learning framework. Finally, model performance was systematically quantified using Accuracy, Loss, Confusion Matrices, and Macro ROC-AUC, with the optimal network weights automatically preserved based on the highest validation accuracy achieved during training.

### 4.1. Convergence curves

The training dynamics of the compared models are illustrated in Figure 10 through the validation accuracy and validation loss curves over 100 training epochs. Overall, all models show a rapid improvement during the initial training stage, indicating that the networks are able to effectively learn representative features from the time-series data. However, differences in convergence behavior and stability can be observed among the models.

The FCN model exhibits noticeable fluctuations in validation accuracy, including several sudden drops during training, which suggests unstable learning behavior. The LSTM baseline shows more stable convergence but reaches a relatively lower accuracy level. The ResNet1D model improves stability and achieves higher validation accuracy than both FCN and LSTM. In contrast, the proposed 1D-ConvNeXt-LSTM model consistently maintains the highest validation accuracy throughout most of the training process, indicating stronger feature representation and learning capability.

The validation loss curves further highlight the superiority of the proposed model. While the FCN model presents several significant spikes in loss and the other baseline models converge to higher loss values, the 1D-ConvNeXt-LSTM architecture rapidly decreases the loss in the early epochs and stabilizes at the lowest loss level among all models. This behavior demonstrates more stable training dynamics and better generalization ability. Overall, the results confirm that the proposed 1D-ConvNeXt-LSTM model achieves faster convergence, higher validation accuracy, and lower loss, highlighting its effectiveness and robustness compared with the baseline architectures.

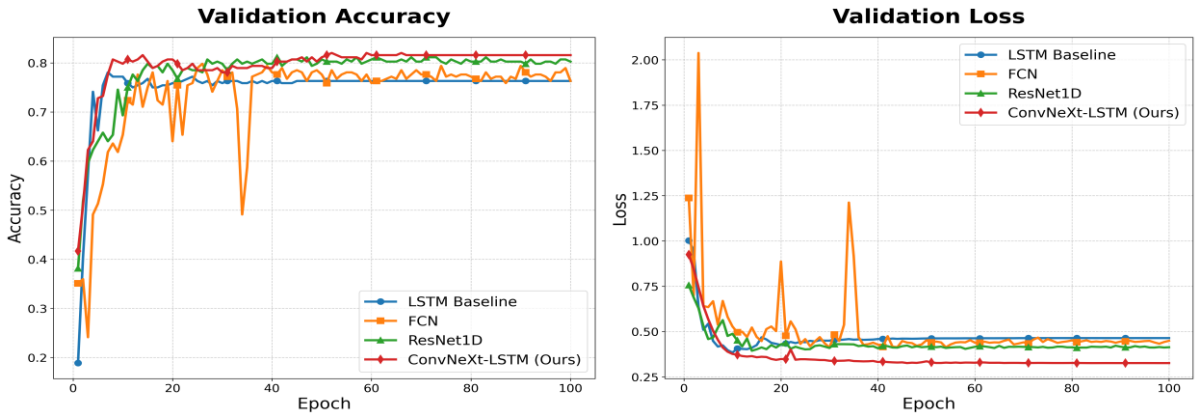


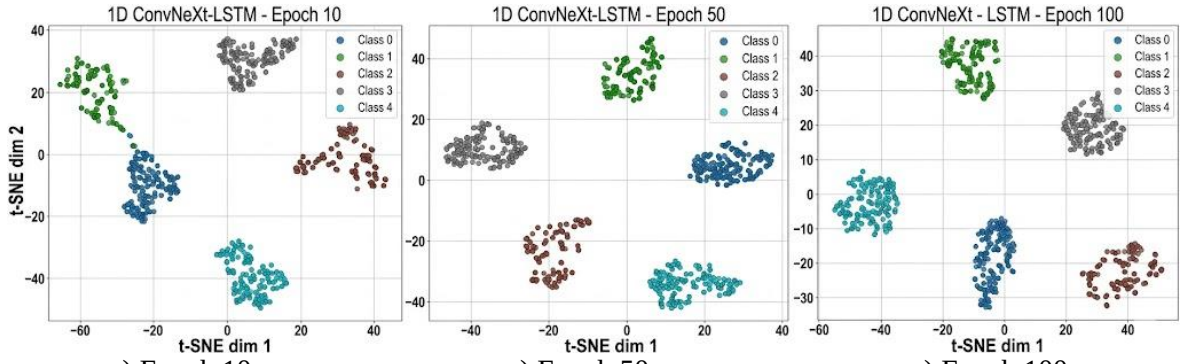

Figure 10. Validation accuracy and loss of different models.

### 4.2. t-distributed Stochastic Neighbor Embedding (t-SNE)

The t-SNE visualization was used to analyze the feature representation learned by the 1D ConvNeXt-LSTM model at different training stages (epoch 10, 50, and 100). The results (Figure 11) illustrate how the learned feature space evolves as the training process progresses.

Overall, the progressive improvement in cluster separation across the training epochs confirms the effectiveness of the 1D ConvNeXt-LSTM architecture in extracting discriminative features from the sensor data, which contributes to reliable structural condition identification and damage detection.

From Epoch 10 to Epoch 100, the model's ability to separate features improves significantly. Initially, the clusters remain close together and are only partially distinct. However, as training progresses, samples belonging to the same class form tighter groups, and the margins between different classes widen. By Epoch 100, the clusters are clearly separated with minimal overlap, demonstrating that the model has successfully extracted the core features required to accurately classify structural conditions.




a) Epoch 10 a) Epoch 50 a) Epoch 100

Figure 11. t-SNE visualization of 1D-ConvNeXt-LSTM features.

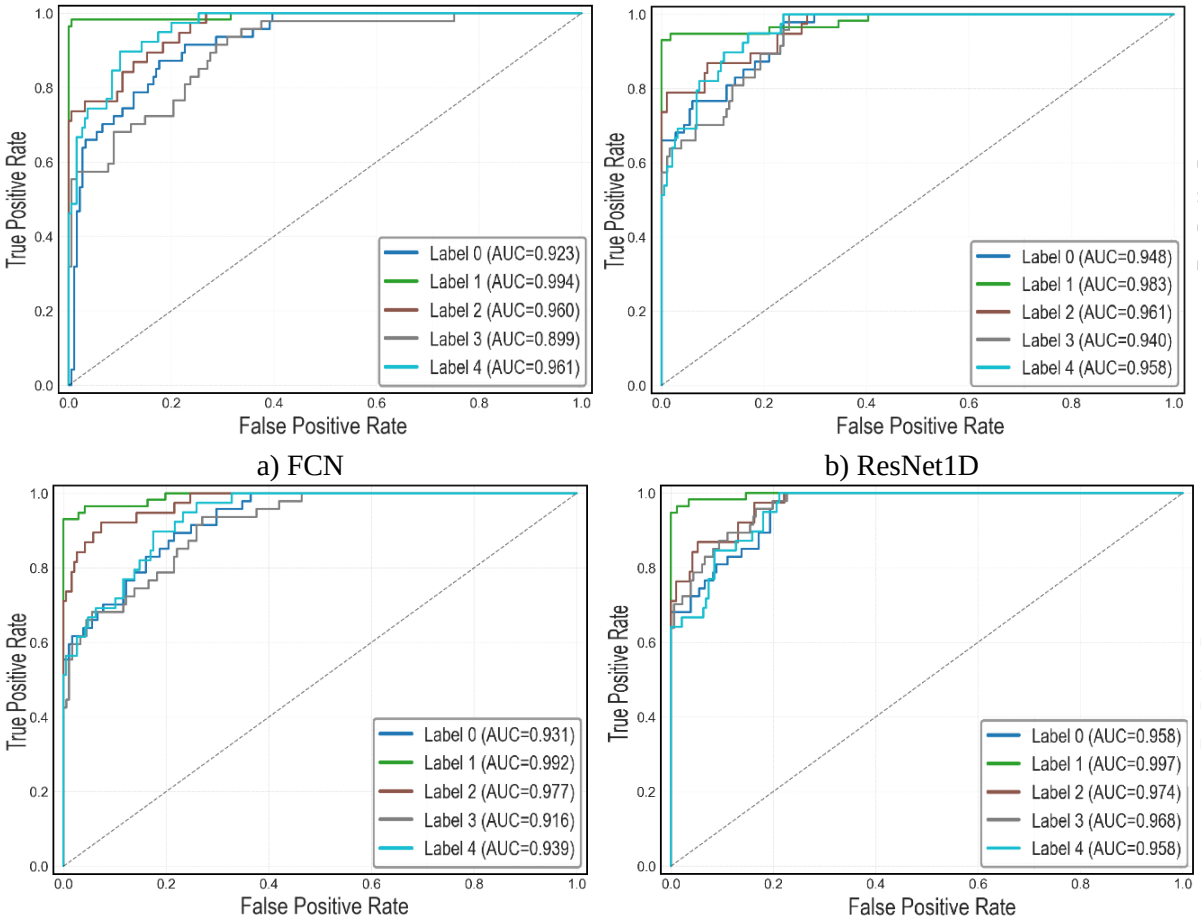
### 4.3. ROC-AUC curves

The classification capability of the compared models was further evaluated using Receiver Operating Characteristic (ROC) curves and the corresponding Area Under the Curve (AUC) values on the validation dataset. ROC curves illustrate the trade-off between the true positive rate and false positive rate across different thresholds, while the AUC value provides an overall measure of model discrimination ability. As shown in Figure 12, all models achieve relatively high AUC values, indicating strong capability in distinguishing structural states from multi-sensor time-series signals.

Among the baseline models, FCN, ResNet1D, and LSTM achieved macro AUC values of 0.947, 0.958, and 0.951, respectively. Although these models demonstrate good classification performance, their ROC curves reveal slight inconsistencies across different classes, particularly for several structural states where the curves deviate further from the ideal top-left region.

The proposed 1D-ConvNeXt-LSTM model achieves the best overall performance with a macro AUC of 0.971, outperforming all baseline architectures. The ROC curves of the proposed model are consistently closer to the upper-left corner across all classes, indicating stronger discrimination capability and more stable performance. This improvement demonstrates that integrating ConvNeXt-based feature extraction with LSTM temporal

modeling enables the model to effectively capture both local signal patterns and long-term temporal dependencies, leading to more accurate and robust structural state classification.




a) FCN b) ResNet1D

c) LSTM d) 1D-ConvNeXt-LSTM

Figure 12. Comparison of ROC curves and Macro AUC scores across different architectures.

### 4.4. Confusion matrix and F1-score

The damage classification effects of each model were compared using confusion matrices. In the confusion matrix (Figure 13), the diagonal elements represent the number of correctly classified samples for the corresponding scenarios. The other elements indicate misclassified samples, and the color depth corresponds to the number of samples.

Overall, all four models accurately recognize Label 1, achieving high True Positive values (ranging from 53 to 56 samples). However, performance discrepancies become evident in more complex states. For the baseline models, misclassification is relatively common: FCN frequently misclassifies Label 3 as Label 0 (13 samples), whereas LSTM struggles with Label 4, misidentifying it as Label 0 (8 samples). The ResNet1D model shows improvement in predicting Label 4 but exhibits notable confusion between Label 3 and Label 4 (10 samples of Label 3 misclassified as Label 4).

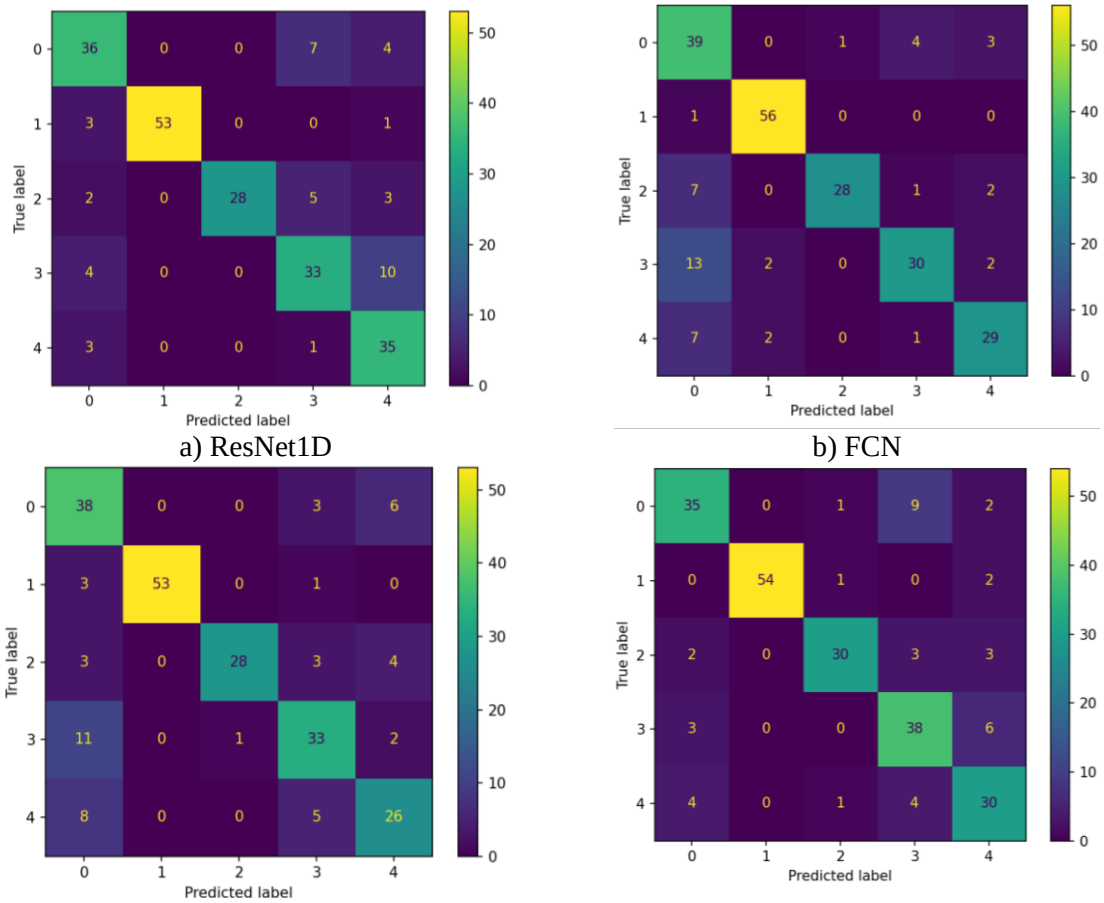



Figure 13. Comparisons of confusion matrices across 4 models.

In contrast, the proposed 1D-ConvNeXt-LSTM architecture demonstrates superior and most consistent performance across all categories. It achieves the highest number of correct predictions in challenging states, specifically Label 2 (30 samples) and Label 3 (38 samples), while significantly minimizing dispersed misclassifications. Although there is minor confusion (9 samples of Label 0 predicted as Label 3), the concentration of values along the main diagonal is the highest and most balanced among all models.

Table 2. Quantitative comparison of classification results.

| Model | Metrics | Label | | | | | Overall accuracy |
|---|---|---|---|---|---|---|---|
| | | 00 | 01 | 02 | 03 | 04 | |
| ResNet1D | Recall | 0.766 | 0.93 | 0.737 | 0.702 | 0.897 | 0.811 |
| | Precision | 0.750 | 1.000 | 1.000 | 0.717 | 0.660 | |
| | F1-score | 0.758 | 0.964 | 0.848 | 0.709 | 0.759 | |
| FCN | Recall | 0.83 | 0.982 | 0.737 | 0.638 | 0.744 | 0.798 |
| | Precision | 0.582 | 0.933 | 0.966 | 0.833 | 0.806 | |
| | F1-score | 0.684 | 0.957 | 0.835 | 0.723 | 0.774 | |
| LSTM | Recall | 0.809 | 0.930 | 0.737 | 0.702 | 0.667 | 0.781 |
| | Precision | 0.603 | 1.000 | 0.966 | 0.733 | 0.684 | |
| | F1-score | 0.691 | 0.964 | 0.835 | 0.717 | 0.675 | |
| **Our** | Recall | 0.745 | 0.947 | 0.789 | 0.809 | 0.769 | 0.820 |
| | Precision | 0.795 | 1.000 | 0.909 | 0.704 | 0.698 | |
| | F1-score | 0.769 | 0.973 | 0.845 | 0.753 | 0.732 | |

The classification performance of the proposed model was further evaluated using the F1-score, which balances precision and recall and is particularly suitable for multi-class classification problems.

The quantitative results in Table 2 show that the proposed 1D-ConvNeXt-LSTM model achieves the best overall performance among all compared methods, obtaining the highest overall accuracy of 0.820. In terms of class-wise performance, the proposed model also achieves the highest F1-scores for Class 0 (0.769), Class 1 (0.973), and Class 3 (0.753) compared with the baseline models. Although ResNet1D slightly outperforms the proposed model in Class 2 with an F1-score of 0.848, the proposed architecture still maintains competitive performance with an F1-score of 0.845 for this class. This can be attributed to the high-intensity localized nature of the 30kg load in this scenario, which produces distinct spectral signatures that a pure convolutional network can efficiently map without the need for additional temporal modeling. However, the 1D-ConvNeXt-LSTM model achieves a higher macro F1-score (0.814) and superior overall accuracy, demonstrating better robustness for diverse and complex structural conditions where both local and long-range features are essential.

## 5. CONCLUSION

To address the limitations of standalone models in extracting both detailed local features and long-range temporal dependencies in SHM, this study proposed a novel hybrid architecture, 1D-ConvNeXt-LSTM, for structural damage detection. Compared with traditional baseline or benchmark networks, the proposed framework achieved significant improvements in all key indicators of damage classification. By integrating ConvNeXt-based feature extraction with LSTM temporal modeling, the model can effectively capture multi-scale features and long-term temporal dynamics in the vibration signals of a cable-stayed bridge.

The 1D-ConvNeXt-LSTM model achieved the highest classification performance, obtaining a macro-AUC of 0.971 and a macro F1-score of 0.814, significantly outperforming all baseline architectures. Its ROC curves were consistently closer to the upper-left corner across all classes, demonstrating stronger discrimination capability and more robust structural state classification.

Although the proposed model achieved high classification performance, this study still has certain limitations. Future research will focus on improving noise robustness, verifying the generalization ability of the method on full-scale bridges, and exploring unsupervised or semi-supervised learning approaches.

## ACKNOWLEDGMENT

This research is funded by University of Transport and Communications (UTC) under grant number T2025-CT-008TD

## REFERENCES

[1]. C. R. Farrar, K. Worden, An introduction to structural health monitoring, Philos. Trans. R. Soc. Math. Phys. Eng. Sci., 365 (2007) 303-315. https://doi.org/10.1098/rsta.2006.1928
[2]. D. N. L. Minh, N. H. Xuan, T. V. Manh, B. N. K. Ngoc, Detection of damage in steel truss bridges

using a hybrid 1DCNN–BIGRU model and time-series data augmentation techniques, Transport and Communications Science Journal, 76 (2025) 1281-1295. https://doi.org/10.47869/tcsj.76.9.11
[3]. A. B. Noel, A. Abdaoui, T. Elfouly, M. H. Ahmed, A. Badawy, M. S. Shehata, Structural Health Monitoring Using Wireless Sensor Networks: A Comprehensive Survey, IEEE Commun. Surv. Tutor., 19 (2017) 1403-1423. https://doi.org/10.1109/COMST.2017.2691551
[4]. B. F. Spencer Jr, T. Nagayama, J. A. Rice, Decentralized structural health monitoring using smart sensors, Proceedings of SPIE, the International Society for Optical Engineering/Proceedings of SPIE, 6932 (2008) 693202. https://doi.org/10.1117/12.791077
[5]. B. Peeters, G. De Roeck, One-year monitoring of the Z24-Bridge: Environmental effects versus damage events, Earthquake Engineering and Structural Dynamics, (2001). https://doi.org/10.1002/1096-9845(200102)30:2%3C149::AID-EQE1%3E3.0.CO;2-Z
[6]. A. Tjirkallis, A. Kyprianou, Damage detection under varying environmental and operational conditions using Wavelet Transform Modulus Maxima decay lines similarity, Mechanical Systems and Signal Processing, 66-67 (2015) 282-297. https://doi.org/10.1016/j.ymssp.2015.04.008
[7]. S. Doebling, C. Farrar, M. Prime, A Summary Review of Vibration-Based Damage Identification Methods, Shock Vib. Dig., 30 (1998) 91-105. https://doi.org/10.1177/058310249803000201
[8]. M. Khanahmadi, B. Mirzaei, B. Dezhkam, O. Rezaifar, M. Gholhaki, G. G. Amiri, Vibration-based health monitoring and damage detection in beam-like structures with innovative approaches based on signal processing: A numerical and experimental study, Structures, 68 (2024) 107211. https://doi.org/10.1016/j.istruc.2024.107211
[9]. V. M. Karbhari, L. S.-W. Lee, Vibration-based damage detection techniques for structural health monitoring of civil infrastructure systems, Structural Health Monitoring of Civil Infrastructure Systems, (2009) 177-212. https://doi.org/10.1533/9781845696825.1.177
[10]. Yu. Yu. Shatilov, A. A. Lyapin, A. V. Cherpakov, S. G. Glushko, Vibration based damage detection of bridge structures, E3S Web of Conferences, 402 (2023) 12016. https://doi.org/10.1051/e3sconf/202340212016
[11]. V.-M. Trung, H.-T. Ngoc, M.-T. Duc, L.-B. Phuc, L.-N. Duc, An Effective Damage Detection Approach for a Truss Bridge Using a Hybrid Deep Learning Model, (2025) 91-101. https://doi.org/10.1007/978-981-96-5206-8_10
[12]. H. Tran-Ngoc, Q. Nguyen-Huu, T. Nguyen-Chi, T. Bui-Tien, Enhancing damage detection in truss bridges through structural stiffness reduction using 1DCNN, BiLSTM, and data augmentation techniques, Structures, 68 (2024) 107035. https://doi.org/10.1016/j.istruc.2024.107035
[13]. S.-Y. Kim, M. Mukhiddinov, Data Anomaly Detection for Structural Health Monitoring Based on a Convolutional Neural Network, Sensors, 23 (2023) 8525. https://doi.org/10.3390/s23208525
[14]. N. C. Thi Nguyen, T. M. Vu, Damage detection in structural health monitoring using BiLSTM-1DCNN hybrid network: a case study on a large-scale steel truss bridge, Eng. Comput., 42 (2025) 2226-2242. https://doi.org/10.1108/EC-08-2024-0714
[15]. S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Comput., 9 (1997) 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
[16]. Y. Lee, J. H. Lee, J.-S. Kim, H. Yoon, A Hybrid Approach of Long Short-Term Memory and Machine Learning With Acoustic Emission Sensors for Structural Damage Localization, IEEE Sens. J., 24 (2024) 39529-39539. https://doi.org/10.1109/JSEN.2024.3481411
[17]. N. Lu, Y. Liu, J. Cui, X. Xiao, C. Kang, ConvNeXt-DSAN: An efficient unsupervised learning framework for structural damage identification with monitoring data, Measurement, 261 (2025) 119951. https://doi.org/10.1016/j.measurement.2025.119951