INTERNATIONAL JOURNAL OF INTEGRATED ENGINEERING
ISSN: 2229-838X e-ISSN: 2600-7916

IJIE

Vol. 17 No. 1 (2025) 190-203
https://publisher.uthm.edu.my/ojs/index.php/ijie

A Novel Hybrid Deep Learning-based Approach for Sensor Data Recovery in Structural Health Monitoring

Nguyen Thi Cam Nhung[1]*, Hoang Nguyen Bui[2], Tran Quang Minh[1]

[1] *University of Transport and Communications, Hanoi, VIETNAM*
[2] *State University of New York at Buffalo, New York, USA*

*Corresponding Author: ncnhung@utc.edu.vn
DOI: https://doi.org/10.30880/ijie.2025.17.01.016

Article Info

Received: 20 August 2024
Accepted: 11 December 2024
Available online: 30 April 2025

Keywords

Data recovery, structural health monitoring, hybrid deep learning

Abstract

Structural health monitoring (SHM) systems contribute significantly to ensuring the safety of construction works. However, in reality, data loss often occurs due to many different reasons. A unique hybrid deep learning-based method for recovering sensor data in structural health monitoring (SHM) is presented in this research. The suggested technique accurately reconstructs missing or corrupted sensor data by utilizing the advantages of both Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). While the RNN models time dependencies to recover the missing sequences, the CNN pulls important patterns from the data. The method's great accuracy in recovering sensor data, even under complex circumstances, is proven using case study real-world bridge monitoring data. The steps taken and the analysis of the results are clearly stated in the study. According to the results, the CNN-RNN combination performs better than conventional techniques and provides notable reliability gains for SHM applications. Future studies will try to improve the model even further and investigate how it may be used to a variety of sensor data and structural types.

1. Introduction

Structural Health Monitoring (SHM) plays a crucial role in ensuring the safety, longevity, and performance of bridge structures [1]–[3]. Bridges are exposed to various effects, including traffic loads, environmental conditions, and natural aging, which can lead to wear, damage, or even failure over time [4], [5]. SHM provides continuous monitoring of these structures through sensors that collect real-time data on parameters such as strain, vibration, and displacement. By analyzing this data, engineers can detect early signs of structural degradation, assess the health of the bridge, and implement timely maintenance or repairs [6]. This proactive approach not only enhances the safety of the bridge but also reduces maintenance costs and extends the lifespan of the structure by preventing catastrophic failures. In essence, SHM is a vital tool in modern bridge engineering, contributing to the reliability and sustainability of critical infrastructure [7], [8].

Sensor data plays a central role in SHM systems, particularly for critical infrastructure such as bridges, skyscrapers, and civil infrastructure. Sensors installed on structures are responsible for continuously recording and transmitting information about essential factors such as vibrations, loads, displacements, deformations, temperatures, and pressures [9]. This data provides a comprehensive and detailed view of the current state of the structure, enabling the SHM system to accurately monitor, analyze, and assess the condition of the structure. Sensor data supports strategic decision-making in real-time structural management, helping to enhance operational processes and improve the sustainability of the structure. As technology continues to advance, the collection and analysis of sensor data not only improve the accuracy of SHM systems but also open up numerous potential applications for enhancing safety and efficiency in infrastructure management [10].

However, one of the challenges faced by SHM systems is their strong dependence on continuous and accurate sensor data [11], [12]. When issues such as sensor failures, connection errors, or data loss due to external factors (e.g., harsh environments, signal interference, or conflicts) occur, the monitoring process can be disrupted, significantly reducing the system's reliability and effectiveness. These disruptions affect continuous monitoring and hinder the ability to predict and detect structural issues early, such as cracks, misalignments, or overloads. When data is missing or corrupted, analyzing the actual condition of the structure becomes more difficult, especially in situations that demand high precision to ensure safety [13]. If key parameters such as vibrations, deformations, or pressure are not adequately captured, the SHM system may miss early warning signs of damage or deterioration. This can lead to serious consequences, such as delays in maintenance or even structural failures. In critical situations, incomplete data can also pose potential risks to both the structure and its users. The challenge becomes even greater in complex sensor networks, where the failure of just a few sensors can have a widespread impact on the overall system performance [14]. The lack of information from one sensor not only creates gaps in the data but can also lead to inaccuracies in reconstructing the overall picture of the structure's condition [15]. This is particularly concerning for SHM systems monitoring large and complex structures, such as cable-stayed bridges or high-rise buildings, where sensor data needs to be analyzed within a tightly interconnected framework. Addressing these challenges is crucial to ensuring that SHM systems can provide timely, accurate, and valuable information to maintain the safety and integrity of critical infrastructure.

Many solutions have been researched and proposed to address the issue of missing or corrupted data. These solutions are generally divided into three main categories: model-based approaches, statistical probability methods, and artificial intelligence (AI) applications. The model-based approach is one of the earliest methods applied. The core idea behind this approach is to build a numerical model that accurately reflects the characteristics and behavior of the actual structure. From this model, the missing data can be reconstructed based on the interaction between different parts of the structure and the remaining sensor data. However, this solution often encounters difficulties in practice due to the complexity of input parameters such as loads, boundary conditions, and other external factors. For the model to function accurately, a large amount of information and assumptions about the structure's operating conditions is required. Typically, model-based methods are more commonly used for forward problems—updating the model based on available data—rather than reconstructing missing data [16]–[18].

The statistical probability approach takes a different path, focusing on predicting and restoring data based on mathematical models. Specifically, these methods utilize probabilistic relationships to predict missing or corrupted values in the sensor data set. The advantage of this approach is its flexibility in being applied to various types of data without the need to build a detailed numerical model. However, the major limitation is that the accuracy of this method decreases if too much data is missing or if the relationships between the data points are complex and do not follow simple probabilistic models [19], [20].

The AI solution, particularly deep learning (DL) models, is increasingly popular in handling missing data in SHM. These methods leverage the ability to learn from large datasets, enabling them to understand and predict complex relationships within the data without requiring precise assumptions like model-based or statistical methods. However, a key drawback of this solution is the need for large training datasets and high computational power. Additionally, AI performance heavily depends on the quality of the training data. If the training data is not diverse enough or contains significant noise, the effectiveness of the AI model may fall short of expectations [21], [22].

In recent years, with the rapid development of AI, data recovery methods based on machine learning (ML) algorithms have increasingly demonstrated their superiority across various fields, particularly in SHM. The integration of AI and ML has led to significant breakthroughs, not only in handling large datasets but also in solving complex problems that were previously challenging to address. Thanks to advancements in computing technology, modern ML algorithms now possess powerful computational capabilities, delivering fast and highly accurate results. Moreover, the emergence of DL methods has opened up new opportunities for reconstructing and recovering missing or corrupted sensor data, especially in SHM systems. Deep neural networks have been widely applied to process complex, time-series data patterns. Additionally, DL models can automatically extract features and uncover hidden relationships within the data without human intervention or manual programming. This capability enables SHM systems to recover data more effectively, reducing the risk of missing early warning signs of structural damage or deterioration. Due to these advancements, AI-based solutions, particularly those utilizing deep learning, are becoming indispensable tools in the management and maintenance of critical infrastructure worldwide. In the task of data recovery, AI has been studied as a multi-application. Lei et al. [23] proposed the use of deep convolutional generative adversarial networks to reconstruct lost data. The results of the study demonstrate the method's ability to reconstruct data from different sensors. Oh and colleagues [24] presented a study involving the use of Convolutional neural networks for data recovery in SHM. The data recovered using Oh's method was accurate enough to assess the structural safety. Fan et al. [25] studied and applied densely connected convolutional networks to reconstruct dynamic data of the structure of SHM system. The proposed method has good efficiency, the recovered data fully meets the requirements for in-depth analysis.

In addition, many other studies have addressed the use of AI in regenerating and recovering lost sensor data of SHM systems [26], [27].

Although AI, particularly DL models, has achieved significant success in data reconstruction, there are still considerable limitations. The lack of transparency in AI decision-making processes is a major challenge, as these algorithms operate like "black boxes," making it difficult to explain or verify the internal steps, especially in high-stakes fields like SHM. AI's generalization capability is also limited; when faced with entirely new data or in complex real-world environments, these models may fail to provide accurate predictions, making reliable and effective data reconstruction more difficult. Another limitation is the high demand for computational resources, as modern AI models require powerful hardware and long training times, increasing costs and hindering large-scale adoption [28].

To overcome some of the limitations associated with standalone DL models and to enhance the accuracy of data recovery, this research proposes a hybrid approach that integrates two DL models for reconstructing faulty data. Specifically, a combination of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) is employed to address the issue. The CNN component is responsible for extracting spatial features from the input data, capturing intricate patterns and local dependencies, while the RNN component focuses on modeling the temporal relationships within the data. By leveraging the strengths of both models, this hybrid method offers a more robust and accurate solution for sensor data reconstruction in complex environments. The integration of CNN and RNN enables the system to not only detect and learn detailed features from the sensor data but also to handle sequential dependencies, making it well-suited for applications like SHM where time-series data plays a crucial role in identifying structural conditions. This combined model has the potential to significantly improve the reliability and performance of data recovery systems, minimizing the impact of sensor faults and missing data. The paper is organized into four parts: Introduction, Proposed method, case study, and conclusion. The proposed method presents the implementation steps and related theories in detail. The research on the real dataset of the Thang Long Bridge will be carried out in the case study section. The main contributions are presented in the conclusion section.

2. Proposed Method

2.1 Convolutional Neural Network (CNN)

CNN [29]–[31] is a type of DL model specifically designed for processing grid-like structured data, such as images or time-series data. CNN operates based on the principle of convolution, meaning it applies filters to the input data to automatically extract important features without human intervention. A CNN consists of several different layers, such as the convolutional layer, pooling layer, and fully connected layer. The convolutional layer is responsible for detecting local features from the data, like edges, corners, or more complex shapes, through the use of small filters. Afterward, pooling layers help reduce the size of the data and the number of parameters, enabling the network to learn more efficiently and minimizing overfitting. The fully connected layers at the end of the network are typically used to make predictions based on the extracted features.

CNN is widely applied across various domains, ranging from image recognition and natural language processing to tasks related to time-series data. Notably, with its ability to automatically extract features from data, CNN has proven highly effective in handling and analyzing complex data types, such as satellite images, medical data, and even sensor data in SHM systems. CNN’s capability to recognize local patterns and synthesize information across multiple levels makes it a powerful and essential tool in today's field of artificial intelligence and machine learning.

The basic formula of convolution in CNN can be represented as follows [32]:

$$Y(i,j) == \sum_{m=1}^{M}\sum_{n=1}^{N} X(i+m, j+n).\mathrm{W}(m,n) + b \tag{1}$$

where, *Y(i, j)* is the output of the convolution at position *(i, j); X(i+m, j+n)* is the input value at position *(i+m, j+n)*; *W(m, n)* is the weight of the filter at position *(m, n); b* is the bias term, which is added to the result after the dot product; *M,N* are the dimensions of the filter. This formula essentially computes the weighted sum of the input region covered by the filter and the bias, which forms the basis for feature extraction in CNNs.

2.2 Recurrent Neural Networks (RNN)

RNN [33], [34] is a specialized deep learning model designed to process sequential data, such as time series, text, or audio. Unlike traditional neural network models, RNN has the ability to retain information from previous time steps due to a feedback loop mechanism within the network. This allows RNN to remember and learn from temporal dependencies, enabling the network to predict future values based on previously processed information. In a traditional Recurrent Neural Network (RNN), the main formulas describe how information is computed and passed from one time step to the next. Below are the fundamental formulas of a traditional RNN [35]:

The hidden state h_t at time step t is calculated based on the input x_t at that time and the hidden state h_{t-1} from the previous time step. The formula is as follows:

$$h_t = \phi(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \tag{2}$$

where h_t: Hidden state at time step t; x_t: Input at time step t; W_{xh} *and* W_{hh}: Weight matrix between the input and the hidden state and Weight matrix between the previous hidden state and the current hidden state; b_h: Bias vector; ϕ: Activation function (usually the tanh or ReLU function).
The output y_t at time step t is calculated based on the hidden state h_t:

$$y_t = \sigma(W_{yh}h_t + b_y) \tag{3}$$

where, y_t: Output at time step t; h_t: Hidden state at time step t; W_{hy}: Weight matrix between the hidden state and the output; b_y: Bias vector; σ: Softmax or sigmoid function (depending on whether the problem is classification or regression)

RNN is trained using the Backpropagation Through Time (BPTT) method, where the weights W_{xh}; W_{hh} *and* W_{hy} are updated based on the gradient of the loss function over several previous time steps.

$$\frac{\partial L}{\partial W} = \sum_{t=1}^{T} \frac{\partial L}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial W} \tag{4}$$

The structure of an RNN consists of hidden units, where the output of one-time step becomes the input for the next, creating a chain of repeated calculations. One of the main advantages of RNN is its ability to model data sequences with long-term dependencies. The ability to store and use temporal information makes RNN a powerful tool for processing and analyzing sequential and time-dependent data.

2.3 Combining CNN and RNN for Data Recovery

The combination of CNN and RNN creates a more powerful tool for sensor data recovery, offering significant advantages over using each model individually. First, CNN is employed to extract key spatial features from sensor data, such as hidden patterns or structures within data matrices, thanks to its ability to detect and analyze local features through convolutional layers. These features include essential information about the spatial distribution of the data, helping to reduce the data's size and highlight critical elements that the system needs to analyze. Once the data has been refined through the CNN layers, the next step is to pass it to the RNN layers, where temporal relationships within the extracted features are processed. With its ability to remember and learn from sequential data, RNN plays a crucial role in identifying patterns and trends over time.

This combination results in a system that leverages both the spatial feature analysis capabilities of CNN and the sequential data learning power of RNN. Not only does this enhance data processing efficiency, but it also improves the accuracy of recovering faulty data while minimizing the loss of crucial information. By merging the strengths of both models, the system overcomes the limitations of using each individually, creating a more flexible and robust approach for analyzing and recovering sensor data, particularly in complex and large-scale structural health monitoring systems. Fig.1 shows the data recovery process.

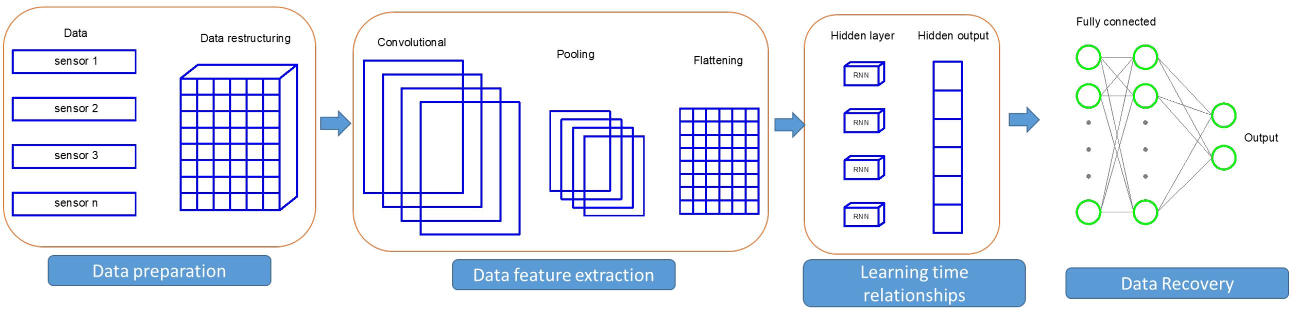


Fig. 1 *Data reconstruction process using CNN-RNN*

After the data is collected from the sensors, the process of preparation for network training begins. Initially, the raw data from the sensors often contains various types of information and is inconsistent, so it needs to be restructured to ensure it meets the input requirements of the proposed model. This restructuring process includes data normalization, noise removal, and rearrangement into the proper format, such as matrix or time-series, to be compatible with the CNN and RNN layers. First, the data will be normalized to the same format with values in the range [0;1]. Next, the data will be rearranged into a matrix and reconstructed into 2-dimensional data. This allows the data to fit the requirements of the proposed model. The input of the proposed model will be the sensors that are working normally, while the output will be the sensors that are faulty.

Once the data is carefully prepared, the CNN layers take on the task of extracting key features. During this stage, the data passes through multiple convolutional and pooling layers, allowing the network to automatically recognize local features, such as patterns or hidden trends in the data. This process not only reduces the size of the data but also highlights the valuable information necessary for the next processing steps.

After the feature extraction step of CNN is completed, the data is flattened in preparation for input into the RNN network. In this step, flattening ensures that spatial features are transformed into sequential time-series data, allowing the RNN network to continue learning and analyzing temporal relationships. With its ability to remember information from previous time steps, the RNN processes and analyzes dependencies over time, enhancing the efficiency of data recovery.

Once RNN has processed the data, fully connected layers are added to the network to complete the reconstruction process. These layers are responsible for combining information from previous layers and generating the final output, which is the recovered or reconstructed sensor data. By using this combined CNN-RNN network, both spatial and temporal features of the sensor data can be effectively handled, thus improving the accuracy and quality of the data recovery process, and minimizing risks and errors in structural health monitoring systems.

3. Case Study

3.1 Introduction to the Case Study – Thang Long Bridge

Thang Long Bridge (Fig. 2) [36] is one of the largest and most prominent steel truss bridges in the capital city of Hanoi, notable not only for its scale but also for its importance within Vietnam's transportation system. This bridge spans the Red River, one of Vietnam's largest and most significant rivers, and plays a strategic role in connecting Hanoi's inner city with northern provinces such as Bac Ninh and Thai Nguyen, as well as key international gateways like Noi Bai Airport and border crossings. The bridge is designed with a special steel truss structure, utilizing two distinct levels.

(a) **(b)**

Fig. 2 *Thang Long Bridge in Vietnam: a. Side view; b. railway floor view*

The upper level is reserved for automobiles, meeting the high demand for road transport in the capital. Its wide roadway design facilitates smooth traffic flow across the bridge, providing crucial support for inter-regional transportation and alleviating congestion on other routes. The lower level of the bridge serves multiple functions, supporting the North-South railway line, one of Vietnam's most vital railways, while also featuring cantilever extensions on both sides to accommodate non-motorized vehicles such as bicycles and motorcycles, helping to efficiently segregate traffic. With its unique design and large scale, the management and maintenance of Thang Long Bridge pose significant challenges, particularly given that the steel structure is susceptible to the harsh weather conditions and environmental factors. However, due to its strategic location and critical role in transportation, Thang Long Bridge remains a symbol of infrastructure development and modernization in Vietnam. Designed according to the 22TCN18-79 standard, the bridge allows the passage of T24 trains and H30-XB80 vehicles. After many years of operation, the bridge has been inspected and assessed for safety in accordance with new standards.

In this study, a vibration data collection was conducted on Thang Long Bridge for research purposes [36]. Accordingly, a measurement grid consisting of 7 points was established and deployed. The data collection task was carried out on one span of the Thang Long Bridge. The sensor installation positions were specified at the truss joints. The design of the measurement points is shown in Fig. 3.

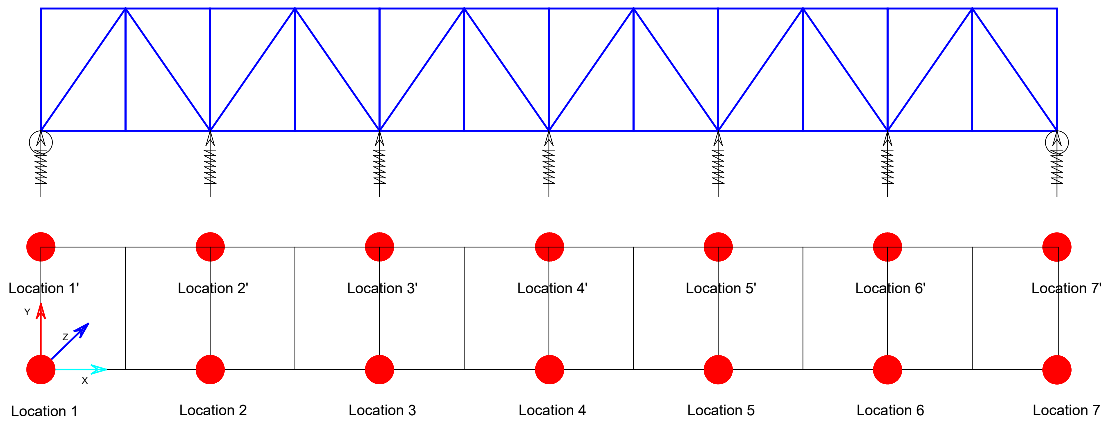


Fig. 3 *Measurement grid at Thang Long bridge*

The equipment used in the data collection campaign includes PCB sensors, an NI data acquisition system, signal transmission cables, a specialized laptop, and other auxiliary devices (Fig. 4). PCB sensors are known for their high sensitivity and ability to measure vibration parameters accurately. In this study, PCB model 393B12 sensor with a sensitivity of 1019.4±10% (mV/(m/s^2)) was used. This type of sensor allows for high-frequency sampling [37]. The NI cDAQ-9178 chassis model was used to install NI 9234 modules specialized for vibration data collection. The NI cDAQ-9178 also manages timing, synchronization, and data transmission between the modules and an external host [38]. Using the NI 9234 module enables the measurement of signals from both Integrated Electronics Piezo-Electric (IEPE) sensors and non-IEPE sensors, such as accelerometers, tachometers, and proximity probes [39]. When paired with NI software, this module provides processing capabilities for condition monitoring, including frequency analysis and vibration status tracking of the structure. The sensors were deployed according to the pre-established plan, securely installed at designated locations on the bridge's joints. Each sensor was carefully connected to the data acquisition system to ensure that the transmitted signal was stable and continuous. Notably, the sampling frequency was set at 1651 Hz, allowing for high-resolution data collection that captures even the smallest fluctuations in the bridge's structure, thus supporting more precise analysis and evaluation.

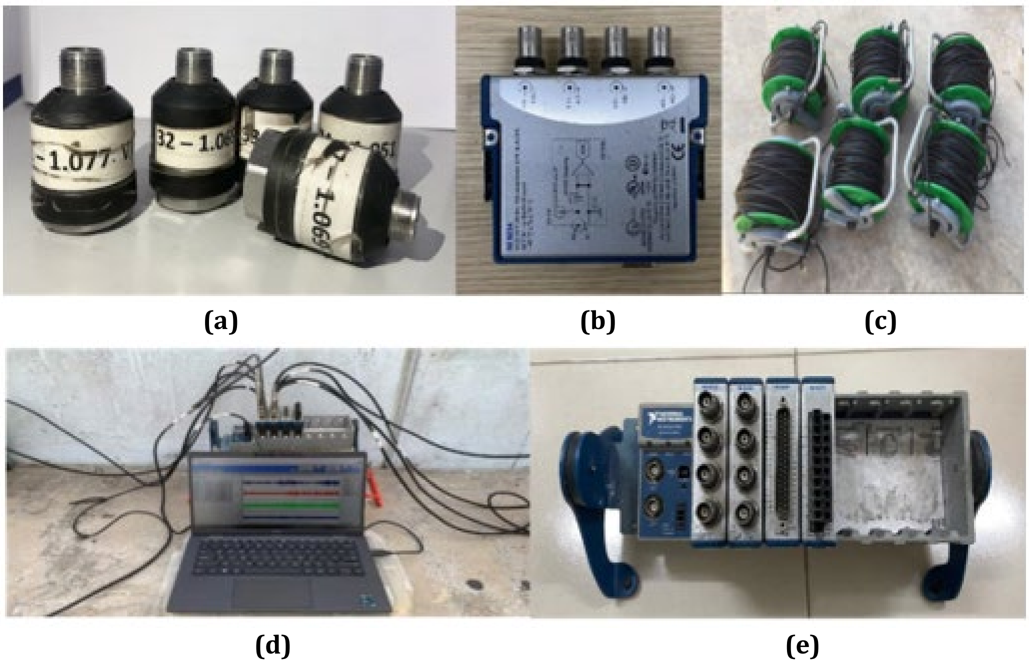


Fig. 4 *Equipment used in data collection campaigns: (a) Sensors model PCB; (b) module NI-9234; (c) Signal transmission cable; (d) Dedicated computer with software; (e) USB Compact-DAQ Chassis*

Fig. 5 presents the data acquisition station at the site and a sensor installation location. On-site, the equipment is securely installed in the designated positions, and a trial run is conducted to check the system's stability. Once the signals from the sensors are confirmed to be stable, the data storage system is activated, and the official data collection process begins. To avoid confusion and ensure smooth data collection, the labels on the sensors and signal transmission cables need to be consistent and clearly marked, allowing the team to easily monitor and manage the operation. After data collection is completed, all the data will undergo preprocessing and temporary on-site checks to ensure the highest quality before being moved to in-depth analysis. This process includes evaluating the correlation between the sensors to confirm that all gathered data is valid and aligns with the actual conditions. If any discrepancies or data issues are detected, the team will make necessary adjustments and perform another round of data collection to ensure the final results meet technical standards and project requirements.

(a) **(b)**

Fig. 5 *(a) Data collection equipment station; (b) The sensor is installed horizontally at location 3*

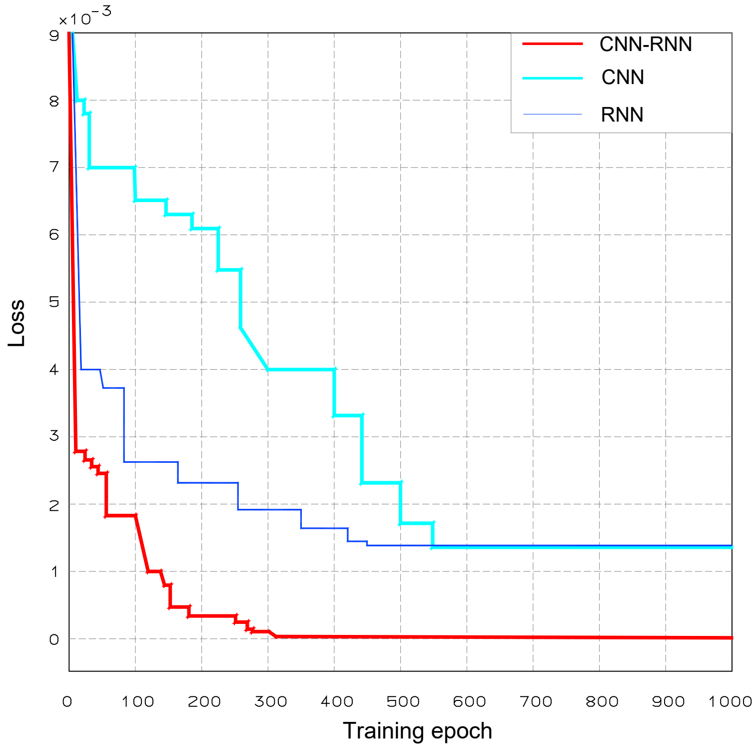
3.2 Single Channel Data Recovery

In the first case study, the process of recovering single-channel data will be conducted. Specifically, one of the sensors will be assumed to have malfunctioned, and its data will be replaced with a value of 0, representing the loss or error in the data collection process. The data from the remaining sensors, which are functioning correctly, will be preserved and used as the input for the CNN-RNN network. These sensors provide critical features and correlations with each other within the structural framework. Afterward, all the data from the active sensors will be input into the designed CNN-RNN network. In this way, the network can predict and reconstruct the missing data from the malfunctioning sensor, ensuring the continuity and completeness of the collected information, which is vital for structural health monitoring and analysis.

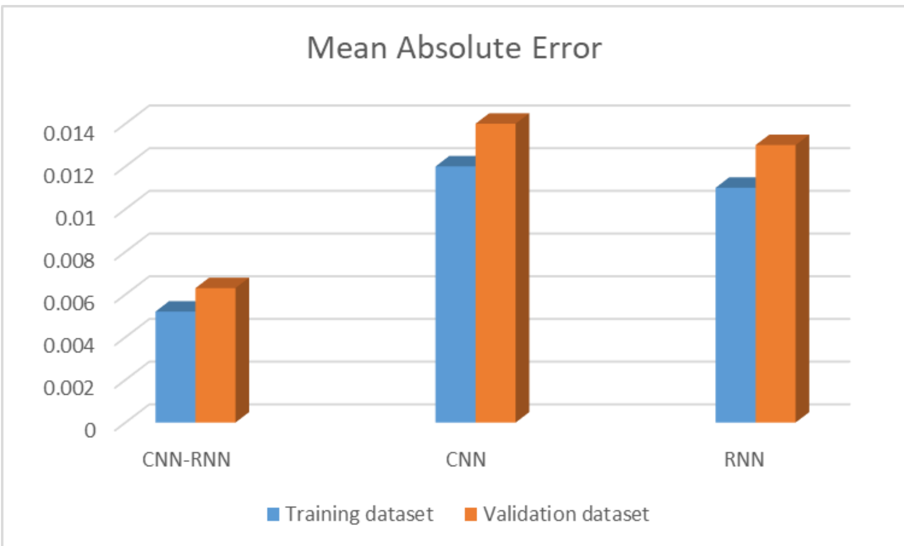
Once the data is restructured, it will be fed into the proposed network for training. The data will be divided at a ratio of 80/20, with 80% used for training the model and the remaining 20% for testing its performance. The network parameters are chosen and fine-tuned based on experience and best practices in the field. In this study, the neural network architecture includes 3 layers of CNN and 2 layers of RNN, specifically designed for the task of data reconstruction. Each CNN layer will have 512 filters, with a kernel size of 5, and the activation function used is ReLU (Rectified Linear Unit). After each CNN layer, a Max-Pooling layer is added to reduce the size and complexity of the data, while helping the network learn more critical features and reducing the risk of overfitting. After passing through the CNN and Max-Pooling layers, the data will be flattened and transferred to the RNN layers. Each RNN layer will have 256 units (neurons) and use the ReLU activation function to process sequence data. Finally, the outputs from the RNN layers will be passed into a Dense layer to produce the final output of the model, which will be used to predict or reconstruct sensor data as required. This neural network structure is designed to leverage the strengths of both CNNs in extracting features from image data and RNNs in processing time series data, with the aim of optimizing the performance of sensor data reconstruction. All data will be trained for a maximum of 1000 epochs. The results of training the CNN-RNN network and some other networks with the same parameters are shown in Fig. 6.

Fig.6 (a) shows the change in loss value over the number of training epochs for three models: CNN-RNN (red), CNN (blue), and RNN (cyan). First, the loss value for all three models gradually decreases over time, indicating that the models are learning and improving their performance. However, the CNN-RNN model (red line) shows the fastest reduction, particularly in the early epochs, demonstrating that it learns more efficiently compared to the other models. CNN-RNN quickly achieves the lowest loss value, highlighting the advantage of combining CNN and RNN architectures. The CNN model (blue line) shows a slower decrease in loss and does not reach as low a loss as CNN-RNN. This suggests that while CNN is good at feature extraction, it is less effective than RNN when handling sequential data or time-dependent relationships. Finally, the RNN model (cyan line) reduces loss more slowly than the other two models, but after a long training period, it also achieves a relatively low loss value. However, compared to CNN-RNN, the RNN takes much longer to converge. Overall, the CNN-RNN model proves to be the most superior in terms of learning speed and final performance, while CNN and RNN converge more slowly and are less effective.

Fig. 6(b) shows the Mean Absolute Error (MAE) values of three models: CNN-RNN, CNN, and RNN on both the training dataset and the validation dataset. From the chart, it is evident that the CNN-RNN model has the lowest MAE on both datasets, indicating the most accurate data reconstruction compared to the other models. Overall, the CNN-RNN model demonstrates the best performance in data reconstruction, while the individual CNN and RNN models perform less effectively. Fig. 7 shows a recovery data segment.



(a) Convergence curve of network training process



(b) Mean absolute error (MAE)

Fig. 6 *Network training results in single-channel data recovery case*

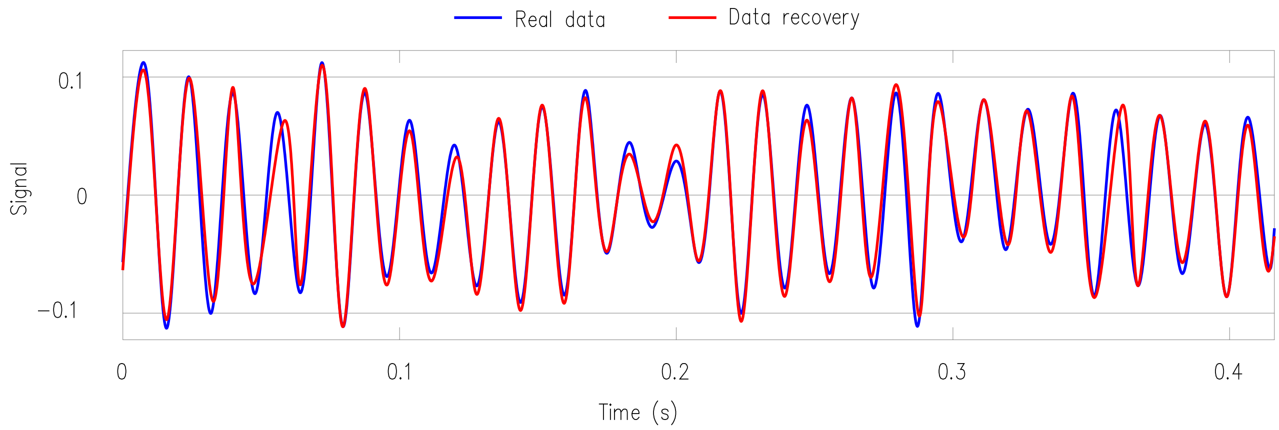


Fig. 7 *Network training results in single-channel data recovery case*

A modal analysis [40] between two datasets, including real data and recovered data, was conducted to assess the model's effectiveness in data reconstruction. This process helps examine the degree of similarity between the recovered data and the original data, thereby evaluating the model's ability to preserve the key characteristics of the original data. The detailed analysis results are presented in Fig. 8 and Table 1.

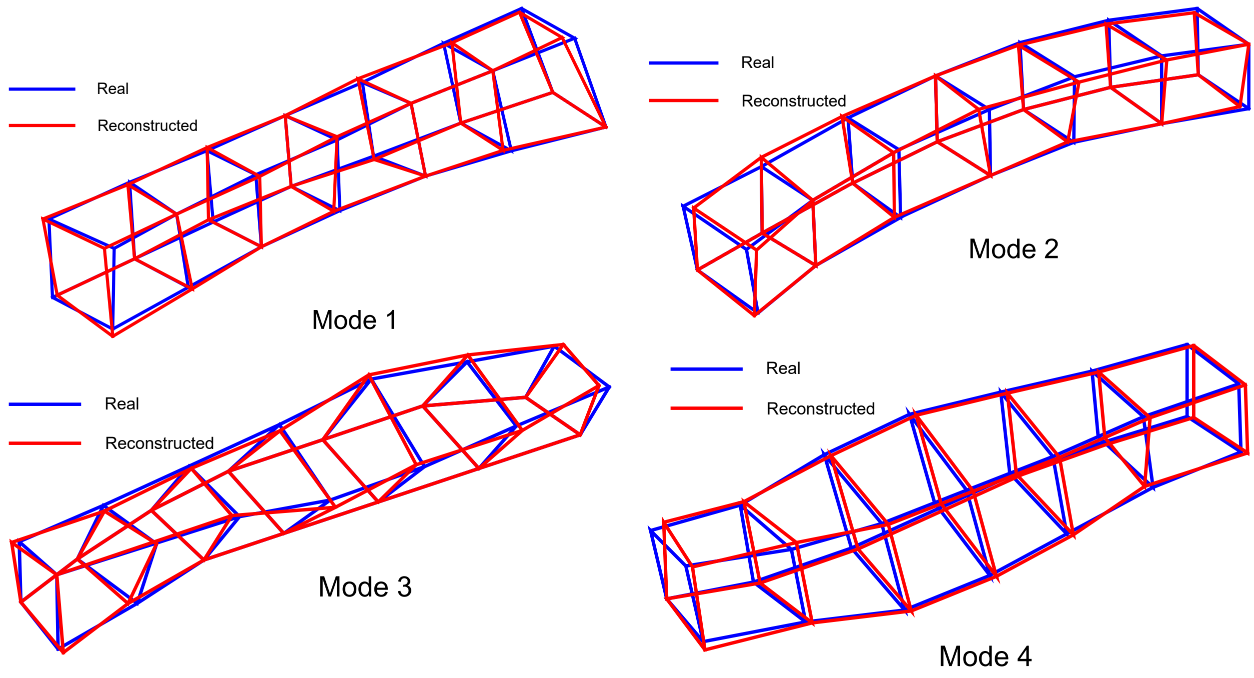


Fig. 8 *Network training results in single-channel data recovery case*

Table 1 *Results of analysis*

Mode	Frequency		Error (%)	MAC
	Real data	Recovery data		
1st	1.05	1.09	3.810	0.966
2nd	1.71	1.61	5.848	0.957
3rd	2.04	2.19	7.353	0.950
4th	2.98	2.81	5.705	0.936

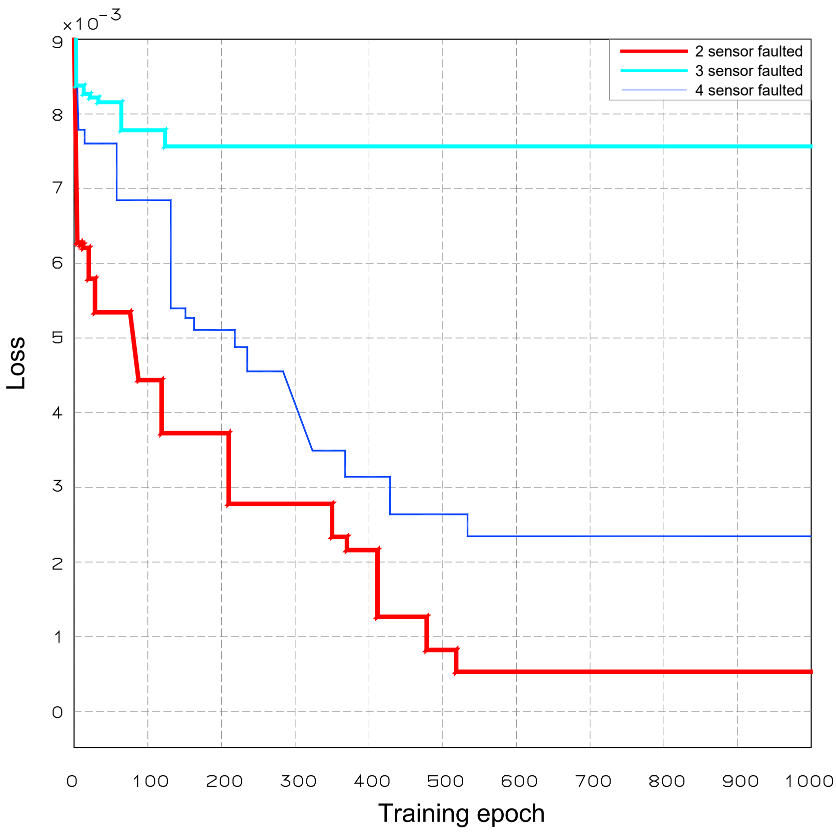
Error = (recovery – real)/real *100%

The modal analysis results between the real and recovery data show a relatively high degree of similarity, although there are some errors. The error percentage is the deviation between the frequency of the recovery data and the actual data. A low error value indicates that the recovered data matches the actual data, meaning the recovery system is accurate. The MAC (Modal Assurance Criterion) value assesses the similarity between the mode shapes of the actual and recovered data. A value close to 1 indicates a high level of similarity, while lower values indicate a lower level of similarity. A high MAC value (close to 1) is desirable because it shows that the recovered data preserves the characteristics of the original mode shape. In the first mode, the frequency of the reconstructed data is 1.09 Hz compared to 1.05 Hz of the real data, with an error of 3.81% and a MAC value of 0.966, indicating a significant similarity between the two datasets. For the second mode, the error increases to 5.848% (real frequency 1.71 Hz and reconstructed 1.61 Hz), with a MAC value of 0.957, still demonstrating a relatively good correlation. However, in the third mode, the error rises to 7.353%, and the MAC value decreases to 0.950, showing that the accuracy of the reconstruction declines. In the fourth mode, the error is 5.705%, and the MAC value is 0.936, which still indicates an acceptable level of similarity. Overall, despite some differences in frequency, the MAC values in all modes are above 0.9, suggesting that the data reconstruction model is capable of accurately reproducing the key characteristics of the real data.

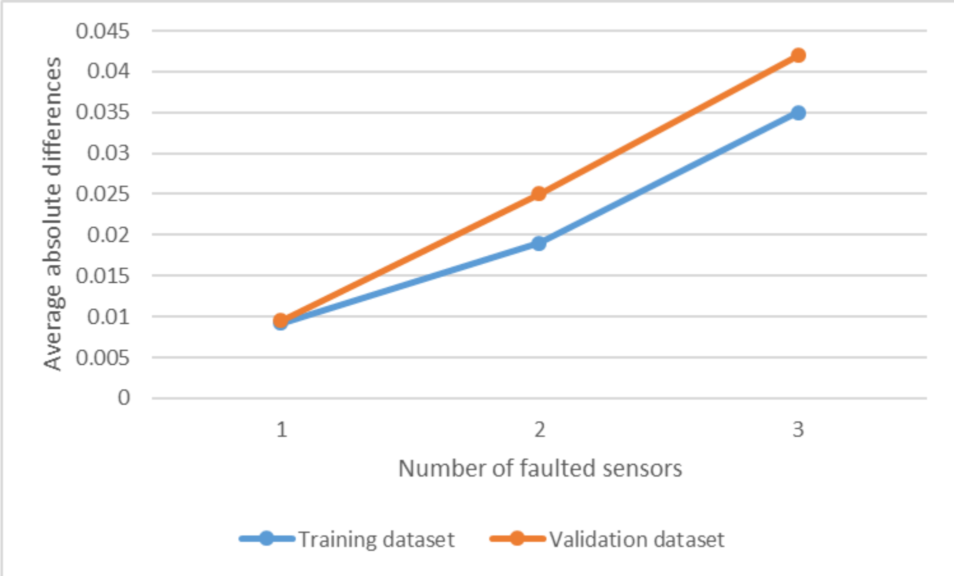
3.3 Multi-channel Data Recovery

To further explore the potential of the proposed CNN-RNN network, the study continued to perform data recovery in a multi-channel scenario, aiming to assess the model's performance under more complex conditions. Specifically, the number of faulty sensors was gradually increased from 2 to 4. The sensors deemed faulty were assigned a value of 0, following the same method applied in the single-channel data recovery case. The goal of this process was to examine whether the model could maintain its effectiveness in reconstructing data when the input information became scarcer, and if so, to what extent.

The configuration parameters of the CNN-RNN network, including the number of filters, kernel size, ReLU activation function, and pooling layers, were kept the same as in the single-channel experiment to ensure consistency and enable direct comparison between the cases. The training and testing process was also performed similarly, with an 80/20 data split for training and validation. The results of the multi-channel data recovery process are displayed in Fig. 9, providing a visual assessment of the model's ability to recover data as the number of faulty sensors increased. This study not only helps define the limitations of the model but also guides the improvement and optimization of the CNN-RNN network for real-world applications, particularly in structural health monitoring systems and sensor data recovery in harsh conditions.



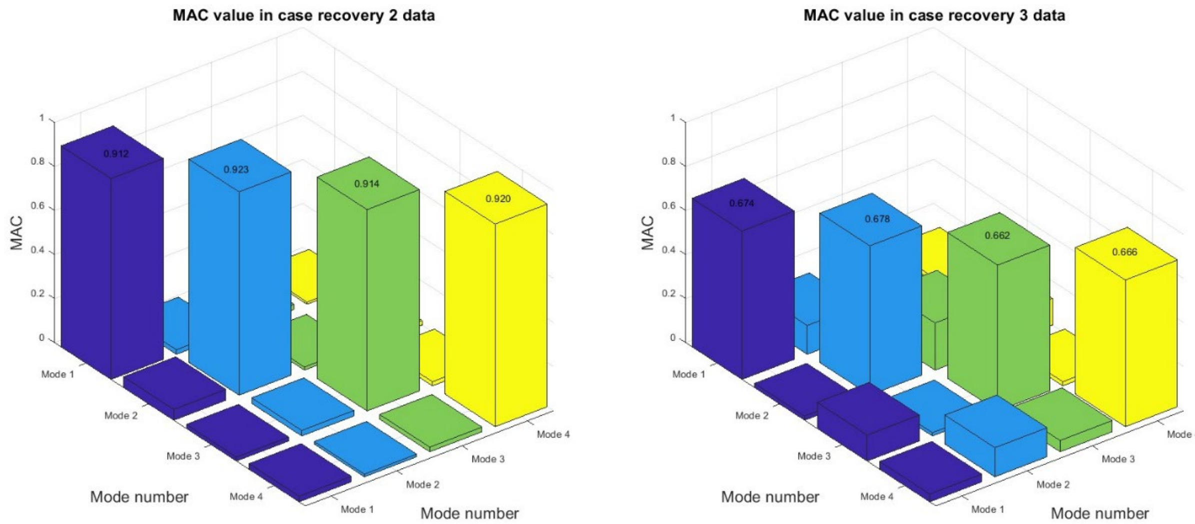
(a) Convergence curve of network training for data loss scenarios



(b) Average absolute differences for data loss scenarios

Fig. 9 *Network training results in Multi-channel data recovery case*

When only two sensors are faulty, the model demonstrates excellent data recovery capabilities, with the loss curve decreasing rapidly and reaching the lowest point among the three cases. In the case of three faulty sensors, the loss reduction process is slower, but the model still achieves a relatively low loss level after a certain number of epochs. However, with four faulty sensors, the loss curve tends to be significantly higher and decreases very slowly, indicating that as the number of faulty sensors increases, the model struggles more with data recovery. This suggests that the CNN-RNN model performs most effectively when the number of faulty sensors is low, and its performance gradually declines as the number of faulty sensors rises. A modal analysis was also conducted on the datasets under different data recovery scenarios. Fig. 10 shows the MAC index obtained after comparing the vibration mode shapes of the datasets.



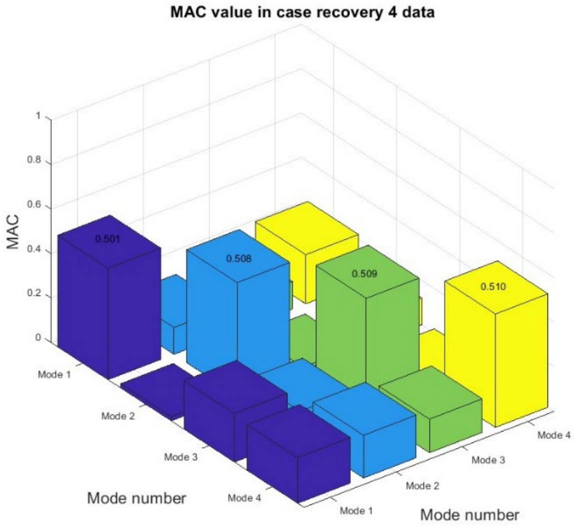


Fig. 10 *MAC values of different data recovery scenarios*

The three charts above display the MAC (Modal Assurance Criterion) values for different data recovery scenarios, as the number of faulty sensors increases from 2 to 4. In the case of recovering 2 faulty sensors, the MAC values are quite high, ranging from 0.912 to 0.920, indicating that the data recovery model performs well and there is a strong correlation between the recovered and real data, especially in the first vibration modes. As the number of faulty sensors increases to 3, the MAC values decrease to a range between 0.674 and 0.666, reflecting a drop in recovery accuracy, although a relatively good correlation is still maintained. However, when the number of faulty sensors increases to 4, the MAC values drop significantly to around 0.501 to 0.510, showing that the model struggles more to reconstruct the original vibration characteristics. Overall, as the number of faulty sensors rises, the accuracy of data recovery decreases significantly, particularly in the higher vibration modes.

4. Conclusions

This study proposes a combination of CNN-RNN models for sensor data recovery in SHM. Through the case study on the dataset from Thang Long Bridge, the effectiveness of the proposed method is evident. The model demonstrated its ability to accurately reconstruct data even in the presence of missing or faulty data. The method has shown great potential in improving the accuracy of recovered data, reducing errors, and providing more precise analyses of the monitored structure's condition. Some key conclusions drawn from this study include:

1. The combined CNN-RNN model effectively utilizes the advantages of both architectures. CNN excels at capturing spatial features and patterns from sensor data, while RNN is proficient in handling temporal sequences and dependencies. By integrating these two models, the proposed method enhances the overall accuracy and robustness of data recovery, making it more suitable for complex SHM tasks that require both spatial and temporal data analysis.
2. The proposed method demonstrated high accuracy in recovering data from sensors, even when one or multiple sensors were faulty. The model outperformed standalone CNN or RNN models.
3. The analysis shows that as the number of faulty sensors increases, the accuracy of the model decreases slightly but still maintains a relatively good level of precision. This confirms the robustness of the model when dealing with complex real-world scenarios.
4. The study results demonstrate that the combination of CNN and RNN holds great potential in SHM applications, particularly in ensuring the integrity and accuracy of data, thereby supporting more efficient maintenance and structural management.
5. Further research will focus on improving the accuracy and increasing the training speed of the proposed network.

Acknowledgement

The authors sincerely thank the University of Transport and Communications and State University of New York at Buffalo for their invaluable support and resources throughout this research.

Conflict of Interest

Authors declare that there is no conflict of interests regarding the publication of the paper.

Author Contribution

The authors confirm contribution to the paper as follows: ***study conception and design:*** *Nguyen Thi Cam Nhung, Tran Quang Minh;* ***data collection:*** *Tran Quang Minh;* ***analysis and interpretation of results:*** *Nguyen Thi Cam Nhung, Hoang Nguyen Bui, Tran Quang Minh;* ***draft manuscript preparation:*** *Nguyen Thi Cam Nhung, Hoang Nguyen Bui, Tran Quang Minh. All authors reviewed the results and approved the final version of the manuscript.*

References

[1] Chacón, R., Casas, J. R., Ramonell, C., Posada, H., Stipanovic, I. & Škarić, S. (2023) Requirements and challenges for infusion of SHM systems within Digital Twin platforms. *Structure and Infrastructure Engineering*, 1–17, https://doi.org/10.1080/15732479.2023.2225486

[2] Matos, J., Fernandes, S., Tran, M. Q., Nguyen, Q. T., Baron, E. & Dang, S. N. (2023) Developing a Comprehensive Quality Control Framework for Roadway Bridge Management: A Case Study Approach Using Key Performance Indicators, *Applied Sciences, 13, no. 13: 7985,* https://doi.org/10.3390/app13137985

[3] Nhung Nguyen, Minh Tran, Helder Sousa, Thuc Ngo & Jose Matos. (2022) Damage detection of structural based on indirect vibration measurement results combined with Artificial Neural Network, *J. Mater. Eng. Struct. « JMES »,* https://revue.ummto.dz/index.php/JMES/article/view/3286

[4] Weifeng Tao, Peihui Lin & Naiyu Wang. (2021) Optimum life-cycle maintenance strategies of deteriorating highway bridges subject to seismic hazard by a hybrid Markov decision process model. *Struct. Saf.* https://doi.org/10.1016/j.strusafe.2020.102042

[5] Mark G Stewart. (2001) Reliability-based assessment of ageing bridges using risk ranking and life cycle cost decision analyses, *Reliab. Eng. Syst. Saf.,* https://doi.org/10.1016/S0951-8320(01)00079-5

[6] Minh Tran Quang, Helder Sousa, Binh Nguyen Duc, Jose Matos, Ana Bento, Tiago Ferradosa & Huan Nguyen (2023) Effect of bridge foundation stiffness on dynamic behavior of bridge structure, *IABSE Congr. New Delhi 2023*, https://doi.org/10.2749/newdelhi.2023.0500

[7] Wang, G. & Ke, J. (2024) Literature Review on the Structural Health Monitoring (SHM) of Sustainable Civil Infrastructure: An Analysis of Influencing Factors in the Implementation, *Buildings*, https://doi.org/10.3390/buildings14020402

[8] Jamadin, A., Abdul Kudus, S., Haziq Ya'akob, A. D., Misnan, M. F. & Mohd Jaini, Z. (2023) Vibration-Based Finite Element Model Analysis on Dynamic Characteristics of Ultra-High Performance Concrete Beam, *Int. J. Integr. Eng.* https://doi.org/10.30880/ijie.2023.15.07.020

[9] A. Rahim, S., Manson, G. & Aziz, M. A. (2021) Data Clustering based on Gaussian Mixture Model and Expectation- Maximization Algorithm for Data-driven Structural Health Monitoring System, *Int. J. Integr. Eng.,* https://doi.org/10.30880/ijie.2021.13.07.020

[10] Mayank, M., Paulo, B.L 7 Ramana, G.V. (2022) Structural health monitoring of civil engineering structures by using the internet of things: A review, *J. Build. Eng.,* https://doi.org/10.1016/j.jobe.2021.103954

[11] Preethichandra DMG, Suntharavadivel TG, Kalutara P, Piyathilaka L & Izhar U. (2023). Influence of Smart Sensors on Structural Health Monitoring Systems and Future Asset Management Practices. *Sensors*. https://doi.org/10.3390/s23198279.

[12] Suslu, B., Ali, F. & Jennions, I.K. (2023) Understanding the Role of Sensor Optimisation in Complex Systems, *Sensors*, https://doi.org/10.3390/s23187819

[13] Minh Q. Tran, Hélder S. Sousa & José C. Matos. (2023) Application of AI Tools in Creating Datasets from a Real Data Component for Structural Health Monitoring, https://doi.org/10.1201/9781003306924-9

[14] Vecherin, S.N., Ratmanski, K.D., Hogewood, L., Igor Linkov. (2024) Design of Resilient Sensor Networks Balancing Resilience and Efficiency, *Int J Disaster Risk Sci*., https://doi.org/10.1007/s13753-024-00546-w

[15] Jianhao Gao et al. (2021) Unsupervised missing information reconstruction for single remote sensing image with Deep Code Regression, *Int. J. Appl. Earth Obs. Geoinformation*, https://doi.org/10.1016/j.jag.2021.102599

[16] Minh Q. Tran, Hélder S. Sousa, José C. Matos., Sérgio Fernandes., Quyen T. Nguyen., & Son N.Dang. (2023) Finite Element Model Updating for Composite Plate Structures Using Particle Swarm Optimization Algorithm, *Appl Sci*., https://doi.org/10.3390/app13137719

[17] Suzana Ereiz, Ivan Duvnjak & Javier Fernando Jiménez-Alonso (2022) Review of finite element model updating methods for structural applications, *Structures*, https://doi.org/10.1016/j.istruc.2022.05.041

[18] J.E. Mottershead & M.I. Friswell. (1993). Model Updating in Structural Dynamics: A Survey, *J. Sound Vib*., https://doi.org/10.1006/jsvi.1993.1340

[19] Jianwei Zhang, Minshui Huang, Neng Wan, Zhihang Deng, Zhongao He, Jin Luo (2024) Missing measurement data recovery methods in structural health monitoring: The state, challenges and case study, *Measurement*, https://doi.org/10.1016/j.measurement.2024.114528
[20] Sun, S., Jiao, S., Hu, Q., Wang, Z., Xia, Z., Ding, Y., & Yi, L. (2023) Missing Structural Health Monitoring Data Recovery Based on Bayesian Matrix Factorization, *Sustainability*, https://doi.org/10.3390/su15042951
[21] Deng, Y., Ju, H., Li, Y., Hu, Y., & Li, A. (2022) Abnormal Data Recovery of Structural Health Monitoring for Ancient City Wall Using Deep Learning Neural Network. *Int. J. Archit. Herit*. https://doi.org/10.1080/15583058.2022.2153234
[22] Ba Panfeng, Zhu Songlin, Chai Hongyu, Liu Caiwei, Wu Pengtao, & Qi Lichang (2024) Structural monitoring data repair based on a long short-term memory neural network, *Sci Rep*., https://doi.org/10.1038/s41598-024-60196-2
[23] Lei, X., Sun, L., & Xia, Y. (2021) Lost data reconstruction for structural health monitoring using deep convolutional generative adversarial networks, *Struct. Health Monit*., https://doi.org/10.1177/1475921720959226
[24] Oh, B., Glisic, B., Kim, Y., & Park, H. (2020), Convolutional neural network–based data recovery method for structural health monitoring, *Struct. Health Monit*., https://doi.org/10.1177/1475921719897571
[25] Gao Fan, Jun Li & Hong Hao. (2020) Dynamic response reconstruction for structural health monitoring using densely connected convolutional networks, *Struct. Health Monit*., https://doi.org/10.1177/1475921720916881
[26] Jiang, H., Wan, C., Yang, K., Ding, Y. & Xue, S. (2022) Continuous missing data imputation with incomplete dataset by generative adversarial networks–based unsupervised learning for long-term bridge health monitoring, *Struct. Health Monit*., https://doi.org/10.1177/14759217211021942
[27] Jiang, K., Han, Q., Du, X. & Ni, P. (2021) Structural dynamic response reconstruction and virtual sensing using a sequence to sequence modeling with attention mechanism, *Autom. Constr*., https://doi.org/10.1016/j.autcon.2021.103895
[28] Zinno, R., Haghshenas, S. S., Guido, G., Rashvand, K., Vitale, A. & Sarhadi, A. (2023) The State of the Art of Artificial Intelligence Approaches and New Technologies in Structural Health Monitoring of Bridges, *Appl Sci.,* https://doi.org/10.3390/app13010097
[29] Zoumana Keita. (2018) An Introduction to Convolutional Neural Networks (CNNs). https://www.datacamp.com/tutorial/introduction-to-convolutional-neural-networks-cnns.
[30] Yamashita, R., Nishio, M., Do, R.K.G. et al. (2018) Convolutional neural networks: an overview and application in radiology, *Insights Imaging*, https://doi.org/10.1007/s13244-018-0639-9
[31] Olga Mierzwa-Sulima. (2020) Convolutional Neural Networks: An Introduction. https://www.appsilon.com/post/convolutional-neural-networks.
[32] Venkatesan, R. & Li, B. (2017) Convolutional Neural Networks in Visual Computing a Concise Guide. https://doi.org/10.4324/9781315154282
[33] Glenn Pietila & Teik C. Lim. (2012). Intelligent systems approaches to product sound quality evaluations – A review, https://doi.org/10.1016/j.apacoust.2012.04.012
[34] Jason Brownlee (2023) An Introduction to Recurrent Neural Networks and the Math That Powers Them. https://machinelearningmastery.com/an-introduction-to-recurrent-neural-networks-and-the-math-that-powers-them/
[35] Sushmita Poudel. (2023) Recurrent Neural Network (RNN) Architecture Explained. https://medium.com/@poudelsushmita878/recurrent-neural-network-rnn-architecture-explained-1d69560541ef
[36] Nguyen Nhung, Hoang Bui, Minh Tran. (2024) Reconstructing Health Monitoring Data of Railway Truss Bridges using One-dimensional Convolutional Neural Networks, *Eng Technol Appl Sci Res*., https://doi.org/10.48084/etasr.7515.
[37] PCB Piezotronics an Amphenol Company (2024) Accelerometer, ICP®, Seismic, Model 393B12. https://www.pcb.com/products?m=393b12
[38] National Instruments NI. (2024) cDAQ-9178, CompactDAQ chassis (8 slot USB). https://www.ni.com/pt-pt/shop/model/cdaq-9178.html?srsltid=AfmBOooB3pjn3xgbMzAw6O9b5cwkhX_job4YNbQ6RpJ71PMiDIfAMwv6.
[39] National Instruments. (2024) NI-9234 Specifications 2023. https://www.ni.com/docs/en-US/bundle/ni-9234-specs/page/specs.html?srsltid=AfmBOopm37_6pQ5538EpTDD1ZN8N77zOgkiV-oM60e5jKCeuC5Ehemzc.
[40] Edwin Reynders, Mattias Schevenels & Guido De Roeck. (2024) Macec - The matlab toolbox for experimental and operational modal analysis. vol. MACEC 3.4. https://bwk.kuleuven.be/bwm/macec/macec.pdf