Li, Zhenkun; Lan, Yifu; Lin, Weiwei

Footbridge damage detection using smartphone-recorded responses of micromobility and convolutional neural networks

Published in:
Automation in Construction

DOI:
10.1016/j.autcon.2024.105587

Published: 02/07/2024

Document Version
Publisher's PDF, also known as Version of record

Automation in Construction 166 (2024) 105587

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Footbridge damage detection using smartphone-recorded responses of micromobility and convolutional neural networks

Zhenkun Li *, Yifu Lan, Weiwei Lin

Department of Civil Engineering, Aalto University, 02150 Espoo, Finland

ARTICLE INFO

Keywords:
Structural health monitoring
Vehicle–bridge interaction
Smartphones
Scooters
Convolutional neural networks

ABSTRACT

This paper presents a footbridge damage detection and classification framework using smartphone-recorded responses of micromobility and deep learning techniques. Time–frequency representations (TFRs) of scooter vibrations are employed to detect and classify footbridge damage severities using a Two-Dimensional Convolutional Neural Network (2D CNN). A One-Dimensional (1D) CNN using scooter frequency spectra was also investigated for comparison. The effectiveness of the method was verified using a numerical model of scooter-footbridge interactions and field tests on a real footbridge. The results indicated that both CNNs were sensitive to footbridge frequency alterations caused by damage in the numerical simulations. Nevertheless, the performance of the 1D CNN experienced a substantial decline in field tests involving stochastic influencing factors, whereas the accuracy of damage classification using the 2D CNN remained high. Finally, reasonable interpretations for the superior performance of the 2D CNN are provided using Shapley Additive Explanations (SHAP) values.

1. Introduction

Structural health monitoring (SHM) of bridges has gained global prominence in recent decades owing to concerns related to aging, deterioration, and the potential risk of collapse [1]. Traditionally, bridge inspection was primarily dependent on vision-based approaches conducted by experienced engineers. However, owing to rapidly increasing human requirements, the scale of bridge structures has reached a new level, rendering manual inspection dangerous or even unachievable [2]. The recent development of sensory devices has highlighted vibration-based methods that provide new perspectives for monitoring the health of structures [3,4]. The core task of SHM is to identify damage to structures and provide information for maintenance, reparation, and early warning of possible failure. This task traditionally involves the installation of different sensors, such as temperature sensors, strain gauges, and accelerometers, directly onto the target bridge (i.e., the direct method) so that data can be continuously collected and analyzed [5,6]. However, several challenges have been identified in practical applications, especially the high cost of sensory networks. Consequently, monitoring systems are typically prioritized for large-scale or essential bridges. Numerous footbridges that significantly facilitate humans' daily lives may not have been properly monitored.

An alternative approach, known as the indirect method, was introduced by Yang et al. [7] in 2004 and subsequently validated experimentally by Lin and Yang [8] in 2005. It was found that the bridge's fundamental frequency was identifiable in the vibrations of the quarter-car model and experimental tractor-trailer system. Since sensors are installed on vehicles rather than bridges, checking, maintaining, and repairing the sensory system becomes easier once faults are noticed. Furthermore, this approach significantly reduced the costs associated with bridge monitoring. Consequently, indirect methods have attracted considerable attention. Numerous studies involving numerical simulations, laboratory experiments, and field tests have been published [9,10].

Contemporary research endeavors on indirect methods have universally focused on identifying bridge frequencies and modal shapes by analyzing vehicle responses. Two key obstacles are typically highlighted: the vehicle's self-information regarding its acceleration and the influence of road roughness. To remove the vehicle's dynamic characteristics from its frequency spectrum, the use of special vehicles with frequencies higher than those of the bridge [11], high-damping vehicles [12], and cutting-edge contact-point (CP) responses have been examined [13,14]. However, specialized vehicles may require custom design and manufacturing in industries and may not be available for regular use. In addition, the back-calculation of CP responses typically requires good estimations of vehicular parameters [15], which can be a notable challenge for engineering applications. The second difficulty is closely associated with road roughness. To eliminate this influence,

* Corresponding author.
E-mail addresses: zhenkun.li@aalto.fi (Z. Li), yifu.lan@aalto.fi (Y. Lan), weiwei.lin@aalto.fi (W. Lin).

https://doi.org/10.1016/j.autcon.2024.105587
Received 10 November 2023; Received in revised form 22 June 2024; Accepted 24 June 2024

researchers have explored various approaches, including connected vehicles [16,17], active excitations on bridges [18], and external excitations or amplifiers on vehicles [19,20]. Nevertheless, employing connected vehicles may be challenging because of practical operational difficulties [21] and issues related to the time synchronization of sensors [22]. Active excitation can influence ongoing traffic; however, additional forces applied to vehicles may not be permissible for commercial cars. Furthermore, bridge frequencies may not be sensitive to local damage, mainly when it occurs on a smaller scale [23]. Alterations in the modal parameters and measured signals can be easily masked by operational and environmental noise. These limitations suggest that more sensitive and noise-tolerant data-driven methods can be utilized for damage detection [24].

Owing to advancements in hardware and computer science, machine-learning techniques have been extensively investigated across various domains in recent decades, including damage detection and classification using indirect methods. In 2014, Cerda et al. [25] introduced a support vector machine (SVM) to classify different bridge damage cases simulated using additional masses on the bridge and modifications to the boundary conditions. The results showed that the detection accuracy could equal or surpass that of bridge accelerations. To increase bridge damage classification efficiency and accuracy, different dimension-reduction techniques, such as stacked auto-encoders (SAE), Mel-frequency cepstral coefficients (MFCCs), and Uniform Manifold Approximation (UMAP), have been verified to be effective in eliminating the influence of environmental noise and highlighting damage-sensitive information [26–28]. In 2019, Mei et al. [29] pioneered the application of MFCCs for indirect bridge health monitoring. The MFCCs were extracted for different vehicle runs, and their statistical characteristics were employed as a reference for bridge damage severity estimation. In 2023, Lan et al. [30] classified different health states of a bridge using raw vehicle accelerations. It was found that compared to existing algorithms, the proposed optimized AdaBoost-linear SVM exhibited the potential to enhance the detection accuracy by 5% to 16.7%. In 2022, Mokalled et al. [31] presented a Bayesian estimation technique for detecting and classifying bridge damage. The study showed that the presence, location, and severity of bridge damage could be accurately detected using high-fidelity simulated data from vehicles.

Deep learning, primarily relying on various types of deep neural networks, possesses a superior ability to extract crucial features compared to traditional machine learning methods. To explore the capabilities of deep learning, Malekjafarian et al. [32] suggested the application of artificial neural networks (ANNs) to detect bridge damage by analyzing vehicle responses. The vehicle positions, speeds, and frequency points were incorporated into the models based on either the acceleration or frequency spectra. Numerical simulation results for a quarter-car model and a simply supported beam model demonstrated that damage could be detected, and benchmarks for assessing damage severity were provided. These models were later extended by Corbally and Malekjafarian [33] to include the influence of temperature when CP responses were employed. Convolutional neural networks (CNNs) are a type of deep learning technique that was initially developed for image classification [34]. They have proven to be highly effective in extracting local features from signals and have been widely used in SHM, including applications such as data anomaly detection and structural damage identification [35–37]. In 2017, Abdeljaber et al. [38] introduced a method using CNNs for real-time damage detection and localization using raw acceleration signals from accelerometers installed on structures. To enhance the effectiveness of damage identification using CNNs, in 2023, Shu et al. [39] combined a modal-updating strategy with One-Dimensional (1D) CNNs. They employed an updated finite element (FE) model to generate training data for the CNN training, which enabled damage localization. Additionally, the FE model was used for damage severity quantification by minimizing an objective function that incorporated modal information from the experiments and numerical models. Researchers exploring indirect methods have also recognized the remarkable capabilities of CNNs. In 2020, Locke et al. [24] successfully classified different levels of bridge damage with an accuracy exceeding 80% using CNNs. Their approach considered factors such as temperature, vehicle speed, and weight, albeit limited to frequency responses within the 3–10 Hz range of the vehicle. In 2022, Hajializadeh [40] proposed a transfer learning-based CNN framework and applied it to railway bridges. The results demonstrate that fine-tuned CNNs can accurately and automatically detect and classify bridge damage, even in the presence of varying speeds, rail irregularities, and operational noise. In 2023, Corbally and Malekjafarian [41] introduced a calibrated vehicle and bridge interaction model for CNN model training, which successfully predicted the bridge damage positions and severity. Although several encouraging findings have been presented in the current literature, three concerns can be noted: (1) A considerable majority of studies that employ deep learning techniques have been validated primarily through numerical simulations or laboratory experiments. However, these approaches have not been validated in the field, particularly when unpredictable factors are involved. (2) Conventional indirect methods are constrained to heavy vehicles and are typically not feasible for footbridges. In addition, road bridges and footbridges can have different structural design patterns, which limits the applicability of the current investigations. However, footbridges, which often span tens or even hundreds of meters, play a significant role in everyday life. Therefore, their safety requires equal monitoring [42,43]. (3) Current studies typically utilize the frequency responses or raw accelerations of vehicles as inputs to machine learning models. Few studies have considered vehicular time–frequency representations (TFRs), which have been verified as superior to 1D signals in structural damage diagnosis [44].

The Internet of Things (IoT) network, which incorporates indirect bridge health monitoring, has emerged as a pivotal constituent [45]. Vehicles or micromobility instrumented with an array of sensors can contribute to this process [46]. Recent investigations by Quqa et al. [47] delved into the feasibility of extracting the modal parameters of footbridges using bicycles equipped with smartphones. The findings indicated that while the bicycle was resting on the footbridge, the fundamental frequency of the footbridge could be extracted well. However, the identification of modal parameters can be significantly hindered by both noise and the actions of the rider when passing the footbridge. Electric scooters, characterized by their convenience and lightweight nature, are widely accessible across Europe and are generally allowable on footbridges. They can be easily driven without pedaling forces. Because of their availability and user-friendly characteristics, scooters have the potential to accelerate the development of smart cities.

This study proposes a deep learning-based framework to detect and classify footbridge damage using smartphone-recorded responses from passing scooters. Two types of CNNs, namely, a 1D CNN utilizing the scooter's frequency spectra and a Two-Dimensional (2D) CNN employing its TFRs, were trained to predict the severity of footbridge damage. Accelerations from scooters with different degrees of freedom (DOFs) were stacked to form multiple input channels for the CNNs. Numerical simulations employing a developed rider-scooter model and field tests on a real footbridge were performed to validate the proposed framework. The remainder of this paper is structured as follows: Section 2 outlines the steps in the proposed deep learning framework and the fundamental concepts of CNNs. Section 3 presents numerical simulations employing the developed 4-DOF scooter and a simply supported footbridge model. Section 4 presents a practical case in which a real scooter and footbridge in the field are employed, and the damage detection results using different CNNs are discussed. Finally, concluding remarks are provided in Section 5.

2. Proposed framework

The footbridge damage-detection framework employed in this study is shown in Fig. 1. Initially, as the rider drove the scooter on the footbridge, its acceleration was recorded using smartphones. The recorded

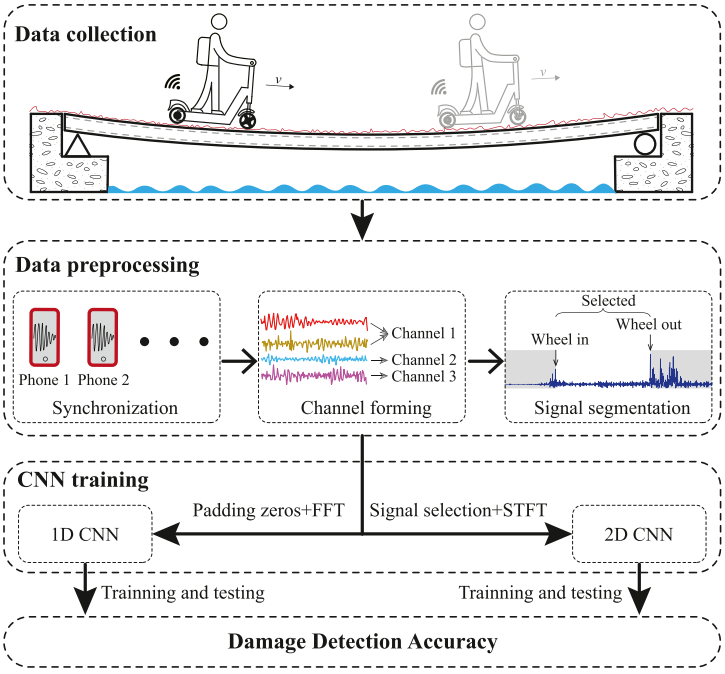


Fig. 1. Framework of the proposed method.

data are then processed to form different channels and subsequently segmented before being fed into different CNNs. Finally, the scooter accelerations are preprocessed by a fast Fourier transform (FFT) to feed the 1D CNN or a short-time Fourier transform (STFT) for the 2D CNN. The dataset was partitioned into training and testing sets to facilitate the calculation of the testing accuracy. The following section introduces the steps and basic concepts of the framework.

2.1. Data collection

In conventional SHM strategies, the placement of sensors is of great importance for capturing damage-sensitive changes in bridge dynamic patterns owing to the numerous DOFs of the structures [48,49]. A salient advantage of using a passing vehicle in the indirect method is that the vehicle itself typically has limited DOFs (when its components are assumed to be rigid, e.g., 4-DOF half-car [50] and 7-DOF full-car models [51]). Theoretically, a vehicle traverses all points on a bridge during its passage, enabling a comprehensive collection of dynamic information from every bridge location. Consequently, this approach provides a new way of monitoring the bridge's health status.

For data collection in this study, smartphones were installed on a scooter to collect vibrations. Initially, the scooter was used to pass the bridge several times when the bridge was in a healthy state (baseline). After the bridge has been in service for months or years and possible damage has occurred, the scooter is utilized again to collect its vibrations when passing the bridge. All scooters running on a healthy or damaged bridge were utilized to train neural networks to predict the health condition of the bridge.

2.2. Data preprocessing

Prior to inputting the scooter data into the CNNs, it is necessary to preprocess the signals to ensure compatibility with different neural networks' input formats. Data preprocessing consists of three steps.

1. Synchronization: In the proposed framework, multiple smartphones can be attached to a scooter to capture its vibrations as it passes through the footbridge. These smartphones can be positioned in different positions on the scooter. However, due to manual operations, data-collecting software cannot be activated and deactivated simultaneously, even with the same sampling frequency. In this study, two smartphones are installed on a single scooter, one attached to the body and the other to the front wheel. We employed the Unix time after collecting accelerations from the two smartphones to synchronize the times.

2. Channel forming. A single smartphone can typically collect acceleration data in six directions: three translational and three rotational. However, when the scooter passes over the footbridge, the most significant signals for analysis are the vertical and pitching accelerations of the scooter. Other vibrations collected in different directions were disregarded to reduce the training parameters of the CNNs. Additionally, to achieve this objective, signal combinations may be necessary to combine acceleration signals from different directions into a single signal. For instance, if a smartphone is not installed perpendicular to the ground, the vertical acceleration must be calculated using signals from two translational directions. Ideally, the selected channels should correspond to the DOFs of a vehicle.

3. Signal segmentation. When a scooter passes over a footbridge, it is expected to maintain a constant speed. Consequently, the scooter must accelerate before reaching the footbridge and decelerate after crossing it. Signal segmentation is required to eliminate any superfluous information (unrelated to the footbridge) during scooter vibrations. This means the collected signals will only be utilized for neural network training when the scooter's wheels are on the footbridge.

Once all of the aforementioned processes have been completed, the CNN input channels will be properly formed. In each channel, only the vibration signals from the scooter on the footbridge were selected for further analysis and training.

2.3. CNN training

CNN is one of the most popular neural networks for extracting damage-sensitive features from signals. In this study, two types of CNNs, 1D and 2D CNNs, based on VGG-16, were developed to extract damage-sensitive features of a footbridge in scooter vibrations. To reduce the number of training parameters, the fully connected layers

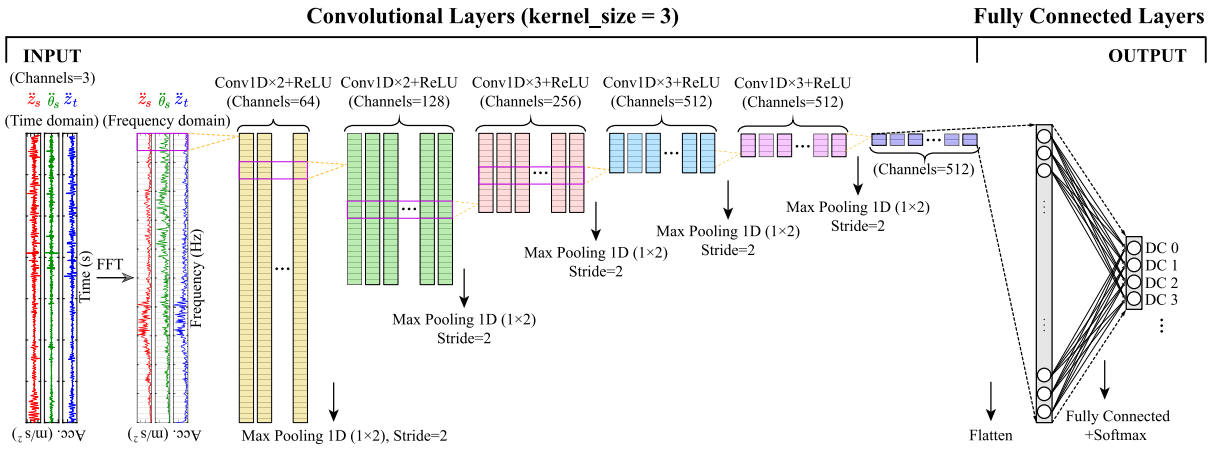


Fig. 2. 1D CNN configuration.

in the original VGG-16 neural network [52] are condensed from four layers to only two layers.

1D CNNs are good at addressing signals recorded by multiple accelerometers at different structural positions, whereas 2D CNNs are specialized in identifying images [53]. In the previous applications of CNNs for monitoring large structures, it was found that a large number of measurement sessions were required. When the sensors were installed on the scooter, the number of measurement points decreased significantly. The inputs for CNNs consist of three key components: $\ddot{z}_s$, representing the vertical acceleration of the scooter body; $\ddot{\theta}_s$, denoting the angular accelerations of the scooter body; and $\ddot{z}_t$, meaning the accelerations of the scooter's front wheel. Experimental measurements can be conducted using two smartphones: one attached to the standing slab captures the body responses of the scooter, and the other attached to the wheel restraint records the accelerations of the front wheel. Even though the scooter is driven at a relatively constant speed when passing the footbridge, its passing time in different runs is challenging to keep strictly the same. To address this issue, when the scooter's frequency spectra were employed for 1D CNN training, the padding zero strategy in the time domain was used to maintain the same FFT resolution for all runs, resulting in the same input dimension for feeding into the neural networks. This approach effectively preserved the information gathered by smartphones. However, when using either the time domain or the TFRs, only a portion of the signals can be utilized to maintain the same input dimensions for all runs. This inevitably leads to a loss of information. Nevertheless, by utilizing TFRs, both the time and instantaneous frequencies can be obtained simultaneously, which is advantageous for identifying footbridge information. In this study, the use of frequency spectra and TFRs was compared using 1D and 2D CNNs.

2.3.1. 1D CNN configuration

Fig. 2 illustrates the configuration of the 1D CNN. A 1D convolution includes several convolutional layers followed by a rectified linear unit (ReLU) activation function. The convolution operation was accomplished using kernels [54]. The kernel size was selected to three with one zero padding at each end of the signal. The maximum pooling kernel size and stride length were both selected as two. Therefore, the data length in one channel in the next layer becomes half the data length in the current layer once the maximum pooling layer is utilized. Finally, the data from all channels were flattened. The softmax layer was used to form different probability values for different classes using Eq. (1),

$$P_{i,n} = \frac{e^{z_{i,n}}}{\sum_{j=1}^{C} e^{z_{j,n}}} \tag{1}$$

where $z_{j,n}$ means the input of the softmax layer, and $P_{i,n}$ represents the probability of being the ith class for the nth sample. The Cross-Entropy (CE) loss was employed in this study [54].

2.3.2. 2D CNN configuration

The architecture of the 2D CNN is shown in Fig. 3, in which TFRs of the passing scooter's vibrations are utilized as input. To ensure that the input has the same dimensions, the signals must be truncated to maintain the same time array after the vibrations of the scooter are collected. Then, transform algorithms such as STFT [44] can be utilized to transfer the time-domain accelerations of the scooter to 2D representations. In this case, the kernel used in each channel is 2D. This study selected the kernel size as three, with one zero padding around the TF image, and ReLU was employed as the activation function. Then, the maximum value in a neighboring 2D area is selected by the max-pooling layers. The flattening process, fully connected layer, softmax layer, and loss function remained consistent with those employed in the 1D CNN.

3. Numerical simulations

3.1. Brief on the VBI model

Fig. 4 shows the rider-scooter model developed by the authors. The model encompasses four DOFs indicated by yellow arrows. When the rider's static mass M_r is 80 kg, its used parameters are an optimized mass $m_r = 1.03M_r$ of 82.4 kg, an optimized rider stiffness $k_r = 1340M_r$ of 107 200 N/m, and an optimized rider damping $c_r = 51.6M_r$ of 4128 N s/m [55]. For the scooter, it owns a body mass m_s of 31.2 kg, body moment of inertia I_s of 101.5 kg m^2, front-wheel mass m_t of 1.17 kg, suspension stiffness k_s of 13 067 N/m, suspension damping c_s of 2000 N s/m, tire stiffness k_{t1}, k_{t2} of 80 000 N/m. In addition, some constants are a_1 = 0.42 m, a_2 = 0.5 m, δ = 0.15 m, ϵ = 0.07 m, and $\varphi = 7\pi/18$ rad [56]. The speed of the scooter is 5.55 m/s when it passes the footbridge. The dynamic equilibrium equations of the rider-scooter model can be given by Eqs. (2)–(5).

$$\mathbf{M}_v \ddot{\boldsymbol{z}}_v^{gc}(t) + \mathbf{C}_v \dot{\boldsymbol{z}}_v^{gc}(t) + \mathbf{K}_v \boldsymbol{z}_v^{gc}(t) = \boldsymbol{p}_v(t) \tag{2}$$

$$\mathbf{C}_v = \begin{bmatrix} c_r + c_s & a_3 c_s & -c_s & -c_r \\ a_1 c_s & a_1 a_3 c_s & -a_1 c_s & 0 \\ -c_s & -a_3 c_s & c_s & 0 \\ -c_r & 0 & 0 & c_r \end{bmatrix} \tag{3}$$

$$\mathbf{K}_v = \begin{bmatrix} k_r + k_s + k_{t2} & a_3 k_s - a_2 k_{t2} & -k_s & -k_r \\ a_1 k_s - a_2 k_{t2} & a_1 a_3 k_s + a_2^2 k_{t2} & -a_1 k_s & 0 \\ -k_s & -a_3 k_s & k_s + k_{t1} & 0 \\ -k_r & 0 & 0 & k_r \end{bmatrix} \tag{4}$$

$$\boldsymbol{p}_v = \left[k_{t1} u_{c2}, a_2 k_{t1} u_{c2}, k_{t2} u_{c1}, 0\right]^T, \boldsymbol{z}_v^{gc}(t) = [z_s^{gc}, \theta_s^{gc}, z_t^{gc}, z_{rd}^{gc}]^T \tag{5}$$

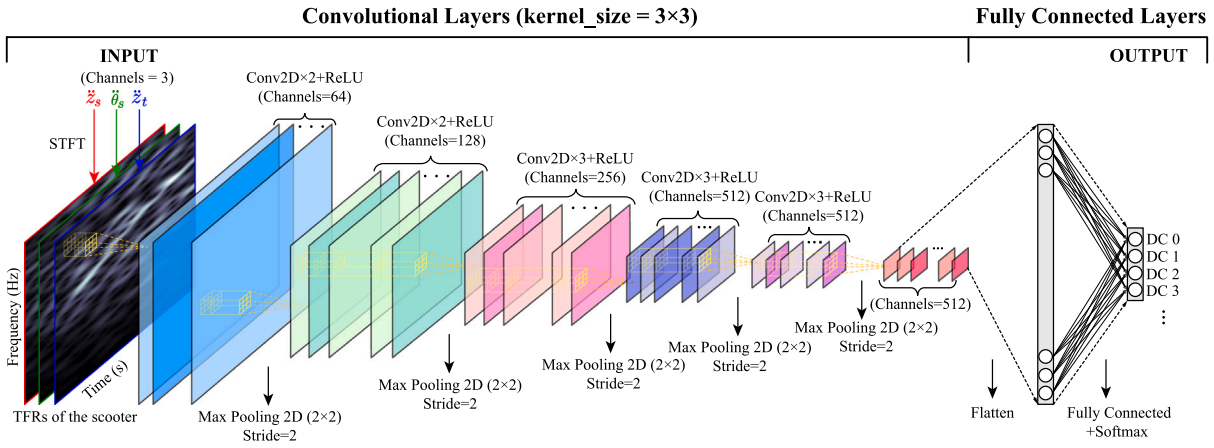


Fig. 3. 2D CNN configuration.

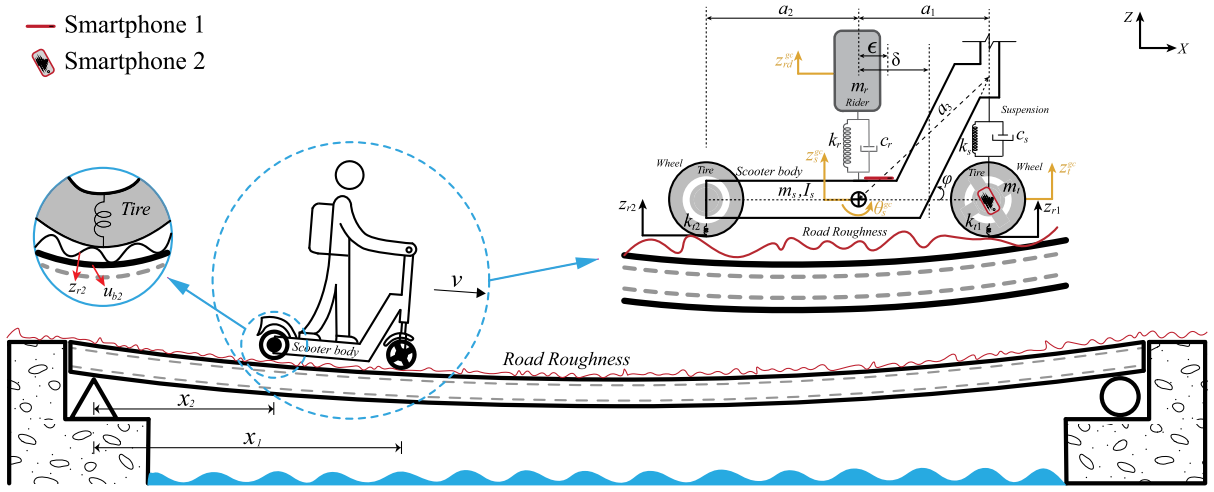


Fig. 4. Numerical simulations of the scooter and footbridge.

where $\mathbf{M}_v = \mathrm{diag}(m_s, I_s, m_t, m_r)$ is the mass matrix, and $\mathbf{C}_v$, and $\mathbf{K}_v$ denote the damping and stiffness matrices of the rider-scooter model. The vectors $\ddot{\boldsymbol{z}}_v^{gc}$, $\dot{\boldsymbol{z}}_v^{gc}$, and $\boldsymbol{z}_v^{gc}$ correspond to the scooter's acceleration, velocity, and displacement, with gc denoting the responses at the scooter's gravity center. The term $\boldsymbol{p}_v$ signifies the external forces acting on the rider-scooter model. The notation u_{ci} signifies the displacement at the ith contact point, which can be determined by the equation $u_{ci} = u_{bi} + z_{ri}, i = 1, 2$, where u_{bi} represents the footbridge's deflection at the ith contact point, and z_{ri} stands for road roughness. The sampling frequency is set at 100 Hz.

It should be noted that capturing the center of gravity of a scooter in practical settings poses difficulties, and the direct recording of its vibrations is not feasible. In the simulations, it was assumed that smartphone 1 was located at a distance of ϵ m from the center of gravity of the scooter body, as shown in Fig. 4. Assuming a rigid scooter body, the simulated vibrations of the scooter are represented by Eq. (6), where the superscript si denotes the simulations.

$$\ddot{z}_s^{si} = \ddot{z}_s^{gc} + \theta_s^{gc} * \epsilon, \ddot{\theta}_s^{si} = \ddot{\theta}_s^{gc}, \ddot{z}_t^{si} = \ddot{z}_t^{gc} \tag{6}$$

The FE model of the simply supported footbridge is shown in Fig. 5. It consists of 20 beam elements, each of which has two nodes with two DOFs (vertical displacement u_b^e and rotation θ_b^e) per node. The equation of motion of the footbridge is given by Eq. (7),

$$\mathbf{M}_b \ddot{\boldsymbol{z}}_b^N(t) + \mathbf{C}_b \dot{\boldsymbol{z}}_b^N(t) + \mathbf{K}_b \boldsymbol{z}_b^N(t) = \boldsymbol{p}_b^N(t) \tag{7}$$

where the mass, damping, and stiffness matrices of the footbridge are represented by $\mathbf{M}_b$, $\mathbf{C}_b$, and $\mathbf{K}_b$, respectively; $\ddot{\boldsymbol{z}}_b^N$, $\dot{\boldsymbol{z}}_b^N$, and $\boldsymbol{z}_b^N$ are the footbridge's accelerations, velocities, and displacements at nodes; $\boldsymbol{p}_b^N$ is the external force vector applied to the nodes. The bridge's parameters are as follows: length $L = 33$ m, element length $l = 1.65$ m, flexural stiffness $E_b I_b = 1.389 \times 10^9$ N m^2 ·, and mass per unit length $\overline{m}_b = 224$ kg/m. For the damping matrix, the Rayleigh damping assumption was utilized with the first two damping ratios $\xi_1 = \xi_2 = 0.02$ [57].

The artificial road roughness was generated according to ISO 8608 [58] with $G_d(n_0) = 4 \times 10^{-6}$ m^3. The road roughness smoothed using a moving average filter (MAF) is shown in Fig. 6. When a scooter is on a bridge, its wheels do not continuously act on the nodes of the footbridge. Once the wheels are not on the nodes, the Hermitian cubic interpolation function can distribute the scooter excitations to the footbridge nodes. The details of road roughness generation and the VBI interaction process can be found in [14]. Furthermore, 10% Gaussian noise was added to the collected scooter accelerations to simulate environmental effects.

3.2. Bridge damage scenarios

A common method for simulating structural damage is to reduce the stiffness of the elements [59,60]. In this study, we examined five simulated damage cases (SDCs), denoted as SDCs 1–5, as shown in Fig. 5, in which the 10th element of the footbridge is damaged. Here, the damage factor μ_{10} is defined as the ratio between the stiffness of the 10th element after damage occurs and that of the undamaged element. SDC 0 with $\mu_{10} = 1.0$ indicates that the bridge is in a healthy state.

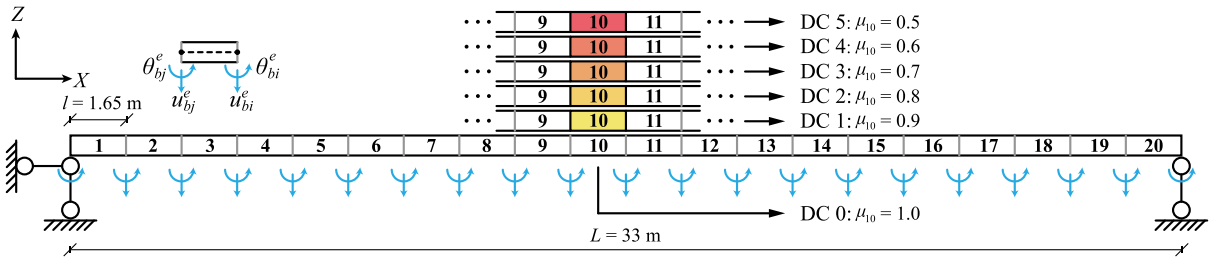


Fig. 5. FE model of the footbridge and damage scenarios.

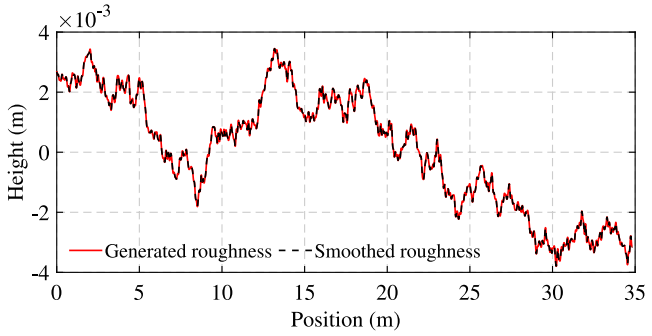


Fig. 6. Road roughness.

Table 1
Footbridge frequencies and runs of different SDCs.

SDCs	SDC 0	SDC 1	SDC 2	SDC 3	SDC 4	SDC 5
f_{b1}^{si}/Hz	3.593	3.573	3.549	3.519	3.480	3.427
f_{b2}^{si}/Hz	14.373	14.369	14.366	14.362	14.362	14.350
f_{b3}^{si}/Hz	32.353	32.173	31.975	31.732	31.732	31.033
Scooter runs	120	60	60	60	60	60

The first three natural frequencies of all SDCs have been provided in Table 1.

The data in Table 1 reveal that the damage employed in various SDCs has only marginal effects on footbridge frequency. Notably, in the case of SDC 1, the alterations in the footbridge's first three frequencies were relatively modest, with changes of only 0.020, 0.004, and 0.180 Hz.

3.3. Damage detection results and discussions

Before employing the CNNs for damage detection, an initial examination of the scooter responses in the healthy state of the footbridge (SDC 0) was conducted. The scooter responses and TFRs are presented in Fig. 7. where $\ddot{z}_s^{si}$, $\ddot{\theta}_s^{si}$, and $\ddot{z}_t^{si}$ denote the simulated scooter accelerations in Eq. (6) with Gaussian noise. Subsequently, FFT was applied to derive the frequency spectra of the scooter accelerations.

The frequency spectra shown in Fig. 7, one can find that when scooter body bounce $\ddot{z}_s^{si}$ and body rotation $\ddot{\theta}_s^{si}$ are utilized, there are peaks around the footbridge's fundamental frequency ($f_{b1}^{si} = 3.593$ Hz). However, as the scooter's second frequency f_{s2}^{si} is quite close to f_{b1}^{si}, it becomes challenging to ascertain whether the peaks at approximately 4 Hz are associated with the scooter or the footbridge. In the simulations, the authors explored the frequencies to which these peaks were linked by deliberately lowering the frequency of the footbridge (e.g., introducing more significant damage, such as $\mu_{10} = 0.1$). The findings revealed that even with a reduction in the fundamental frequency of the footbridge, the peaks persisted at approximately 4 Hz. Thus, it can be inferred that the peaks aligned with the second frequency of the scooter rather than with the fundamental frequency of the footbridge. Nevertheless, researchers face considerable difficulty in differentiating between the frequencies of footbridges and scooters in practical engineering scenarios, mainly when they are very close.

The TRFs of the scooter accelerations were calculated as shown in Fig. 7, in which the STFT is employed with a window length of 100 and an overlapping length of 99. To ensure that the dimensions of the TFRs in all runs were the same, the scooter accelerations in the last 5 s were selected. This temporal segment is preferred because, during these moments, the scooter actively stimulates the footbridge, rendering it more informative about the footbridge's response. From the TFRs, it can be observed that, similar to the frequency spectra, for the scooter body bounce $\ddot{z}_s^{si}$ and body rotation $\ddot{\theta}_s^{si}$, the peaks are more related to the scooter itself instead of the footbridge. In addition, for the scooter's wheel responses, it is difficult to find clear information about the footbridge, predominantly because of significant road roughness and noise contamination.

To detect damage to the footbridge, the proposed framework (Fig. 1) is employed to train the two types of CNNs: 1D and 2D CNNs. In the simulations, the accelerations of the scooter across its three DOFs were readily obtainable and segmentable. Consequently, the initial steps of data collection and preprocessing were omitted. Consequently, the acquired signals can be directly input into different CNNs for training, following zero-padding and FFT for the 1D CNN and signal selection and STFT for the 2D CNN. The number of runs for all SDCs is listed in Table 1.

To train the CNNs, 60 runs originating from SDC 0 were randomly selected to mitigate the data imbalance. Thus, the dataset comprised 360 runs across various SDCs. During the training process, 70% of the runs in the dataset were designated for training, and the remaining runs were allocated for testing. The model training process was conducted within the Python 3.10 environment with Pytorch [61]. The workstation employed at Aalto University was equipped with an Intel Core i9-11900 CPU and 32 GB of RAM. An NVIDIA RTX 3090 graphics card was used to expedite the training procedure.

The configurations of 1D and 2D CNNs are shown in Figs. 2 and 3, respectively. The hyperparameters governing the behaviors of these two neural networks were selected as follows: batch size, 32; optimizer, Adam [62]; learning rate, $1e^{-5}$; weight decay, $1e^{-5}$; loss function, CE loss; and activation function, ReLU. An early stopping strategy was implemented in anticipation of potential overfitting issues, and the number of epochs was set to 400. The accuracy of predicting footbridge damage severity can be calculated using Eq. (8) as follows:

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \times 100\% \tag{8}$$

where TP represents "true positive", FP represents "false positive", TN signifies "true negative", and FN corresponds to "false negative", as shown in [63]. By utilizing the aforementioned hyperparameters and training both CNNs over 400 epochs, the loss and accuracy related to the damage severity classification for both the training and testing phases are plotted in Fig. 8.

It is evident that both CNN architectures achieved a relatively high testing accuracy (>90%). The testing accuracy stabilized at approximately 300 epochs, and the training accuracy reached 100% after the

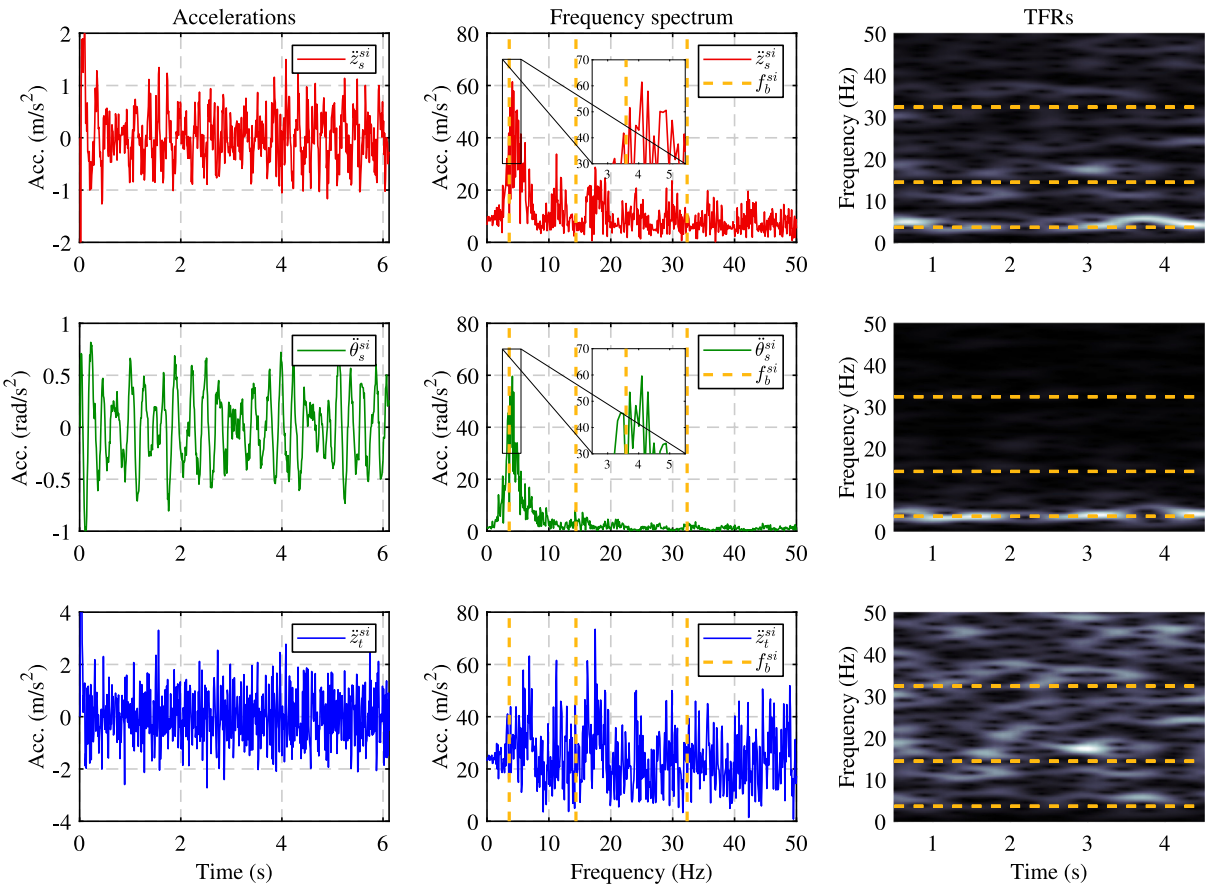
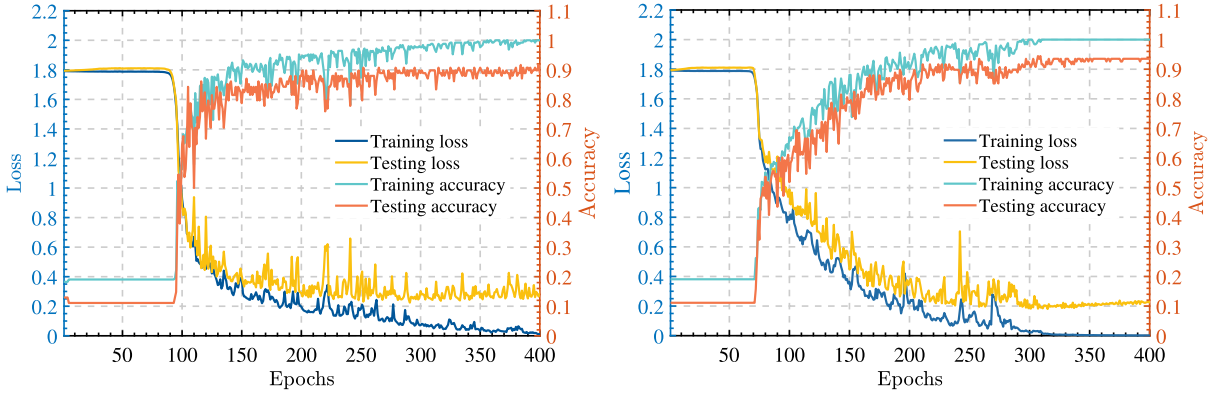


Fig. 7. Responses and TFRs of the scooter in simulations.

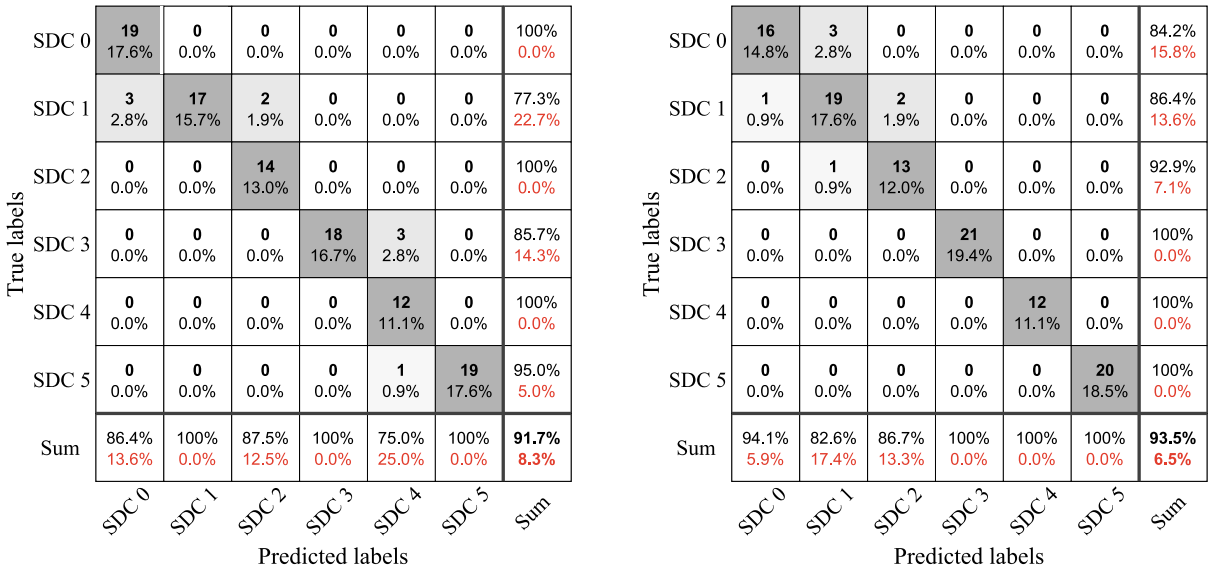


(a) Loss and accuracy when 1D CNN is employed (b) Loss and accuracy when 2D CNN is employed

Fig. 8. Loss and accuracy in numerical simulations using different CNNs.

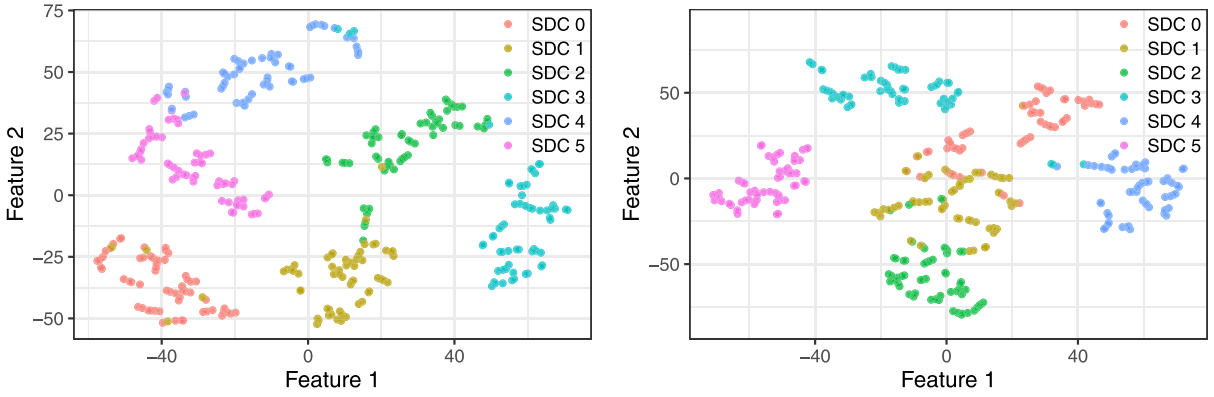
completion of all epochs. No evident overfitting was observed based on the testing loss trends. Moreover, a closer look reveals that the 2D CNN attains slightly superior testing accuracy and lower testing loss when employing the same hyperparameters. The confusion matrices (CMs) illustrating the damage-detection results for all SDCs using both 1D and 2D CNNs are presented in Fig. 9. The results demonstrate that despite a slight reduction in footbridge frequencies in SDC 1–5, both CNN architectures successfully predicted the various health conditions of the footbridge with commendable accuracy, with the 2D CNN exhibiting a marginally better performance.

To demonstrate the feature-extraction capabilities of the employed CNNs, Fig. 10 presents a 2D visualization of the output derived from the convolutional layers using t-Distributed Stochastic Neighbor Embedding (t-SNE). The data points representing the 300 scooter runs at different SDCs were color-coded accordingly. The visualization demonstrates that after the convolutional layers, the input of the fully connected layers is grouped into distinct clusters corresponding to the different SDCs, which verifies the capability of the CNN to extract damage-sensitive features of footbridges from the passing scooter's vibrations.



(a) CM using 1D CNN (b) CM using 2D CNN

Fig. 9. CMs in numerical simulations using different CNNs.



(a) Using 1D CNN (b) Using 2D CNN

Fig. 10. Feature space visualization using t-SNE.

4. Field tests on a real footbridge

4.1. Scooter and footbridge

Field tests were conducted using scooters equipped with smartphones to validate the proposed footbridge damage-detection framework, as shown in Fig. 11. The tests employed two smartphones: an iPhone 12 (smartphone 1) mounted on the scooter's body and an iPhone 8 (smartphone 2) attached to the front wheel. To coincide with the simulations, the scooter was driven to pass through a footbridge with approximately the same road roughness for all the runs. In addition, note that there is a short distance at which the scooter must accelerate to the maximum speed (20 km/h). The scooter can be driven by an electric engine without pedaling forces.

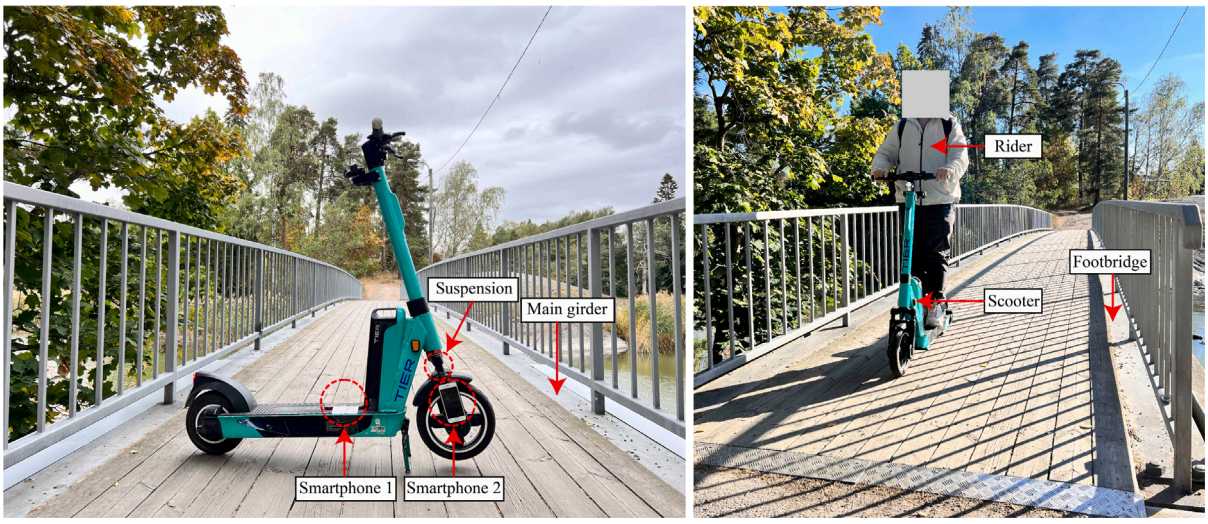
MATLAB Mobile software was employed on the smartphones, as shown in Fig. 12. As the scooter passes the footbridge, the smartphone collects three translational accelerations (X, Y, and Z) and three angular accelerations (RX, RY, and RZ), as indicated by the arrows in Fig. 12. The sampling frequency was set to 100 Hz. The tests were performed in an open area with normal environmental noise.

The footbridge used in the field tests is illustrated in Fig. 13. It was supported at each end and was representative of real-life footbridge structures. In this study, we employed different masses (standing people in the middle of the footbridge) to simulate different experimental damage cases (EDCs), which have been verified as a more practical way to simulate damage to structures [27,64,65]. The specific values for the additional masses were 55, 70, and 60 kg for masses 1, 2, and 3. The various EDCs and the corresponding scooter runs conducted during the field tests are listed in Table 2. The EDC 0 was obtained when there was nobody on the footbridge.

Table 2
Footbridge frequencies and runs of different EDCs.

EDCs	EDC 0	EDC 1	EDC 2	EDC 3
Added mass/kg	0	55	125	185
f_{b1}^{ex}/Hz	4.028	3.979	3.955	3.918
f_{b2}^{ex}/Hz	4.468	4.443	4.431	4.370
f_{b3}^{ex}/Hz	10.486	10.486	10.486	10.437
f_{b4}^{ex}/Hz	11.316	11.304	11.304	11.304
Scooter runs	124	65	63	60

Impulse excitation was applied to the footbridge to investigate the influence of the additional mass. This process is completed by humans jumping on a bridge. One smartphone was placed on the footbridge



(a) Scooter equipped with smartphones (b) Rider and scooter

Fig. 11. Scooter, rider, and smartphones in field tests.

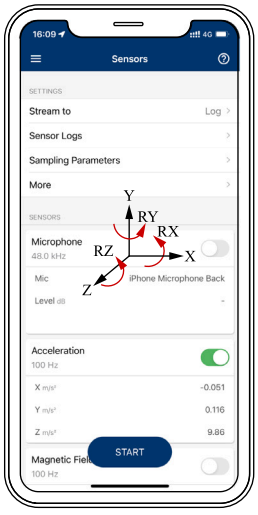


Fig. 12. Smartphone and used software.

(1/4 span) to collect the vibrations. The resulting footbridge accelerations in EDCs 1–3 are shown in Fig. 14(a). It is noteworthy that, in the case of EDC 0, obtaining accurate natural frequencies of the footbridge using the impulse excitation method is unachievable because at least one person applies the excitation force. An alternative approach is to employ the free vibration stage of the footbridge after the scooter passes it, which is introduced in the subsequent section. Fig. 14(b) illustrates the footbridge's frequency spectra in EDCs 1–3. It is evident that increasing the mass can reduce the fundamental frequency of the footbridge. However, similar to the simulations, the decrease in frequency was minimal. All identified frequency values have been documented in Table 2, in which f_{b1}^{ex}-f_{b4}^{ex} (the superscript ex means experiments) represent the first four frequencies of the footbridge. Note that for the first two frequencies f_{b1}^{ex} and f_{b2}^{ex}, there is a decrease, whereas the third and fourth frequencies f_{b3}^{ex} and f_{b4}^{ex} may remain relatively unchanged across the different EDCs.

4.2. Data collection and preprocessing

In the context of the previously described experimental setup, the acceleration of the scooter as it passes over the footbridge can be collected using the two smartphones. However, achieving perfect synchronization in the initiation of data recording is challenging, owing to the use of different smartphones for data collection. In order to address this issue, synchronization is achieved by recording Unix time with a resolution of 0.001 s on smartphones. The details of the synchronization strategy are as follows: The starting timestamp is determined by selecting the maximum value between the starting timestamps of the two signals recorded by the two smartphones. Similarly, the ending timestamp was obtained by selecting the minimum value between the end timestamps of the two signals. The samples between the starting and ending timestamps may not be identical due to sampling time errors. In general, the erroneous timestamp was 0.001 s shorter than the expected sampling timestamp. For the two smartphones used (iPhone 8 and iPhone 12), it was observed that there was typically a difference of one sample within a 10-s interval. In this study, the vibrations during the scooter's passage on the footbridge were typically within 10 s. Therefore, the final synchronization error was approximately 0.01 s for every run. To address this, based on the timestamps recorded by iPhone 12, linear interpolation was employed to obtain the interpolated values of the responses recorded by iPhone 8. By implementing this strategy, we can obtain two synchronized signals of equal length.

Furthermore, to make the series of collected signals correspond to the scooter's DOFs, the accelerations collected by smartphones must be preprocessed using Eq. (9).

$$\ddot{z}_s^{ex} = \ddot{z}_{Z1}, \ddot{\theta}_s^{ex} = \ddot{z}_{RX1}, \ddot{z}_t^{ex} = (\ddot{z}_{X2}^2 + \ddot{z}_{Y2}^2)^{1/2} \tag{9}$$

where $\ddot{z}_{Z1}$ is the translational acceleration collected by smartphone 1 along Z axis, $\ddot{z}_{RX1}$ denotes the angular accelerations along RX of smartphone 1, $\ddot{z}_{X2}$ and $\ddot{z}_{Y2}$ are the translational accelerations collected by smartphone 2 in the X and Y directions, respectively. Therefore, the accelerations $\ddot{z}_s^{ex}$, $\ddot{\theta}_s^{ex}$, and $\ddot{z}_t^{ex}$ corresponding to the three DOFs of the scooter can be formed. The derived acceleration signals were organized into different input channels to train the CNNs in the proposed framework.

Clipping the collected accelerations is essential in removing unrelated vibrations from the scooter's recorded data, ensuring that only the relevant information when the scooter is on the footbridge is used for analysis. Fig. 15 has plotted a series of wheel accelerations when the scooter passes the footbridge. Signal peaks occur when the scooter enters and leaves the footbridge, which can be employed to clip the acceleration when the scooter is on the footbridge.

To compare the direct and indirect methods, footbridge vibrations (at 1/4 span), including the forced and free vibration stages, were

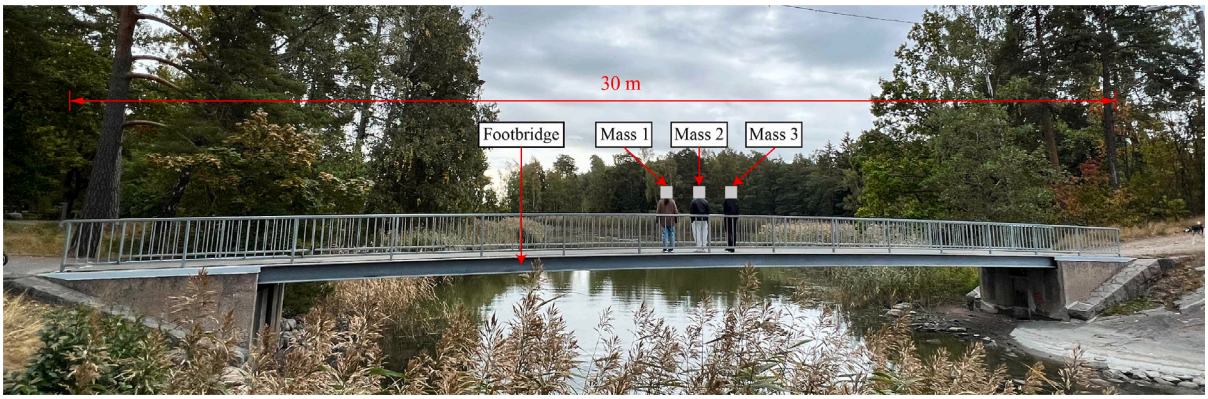


Fig. 13. A footbridge with additional masses.

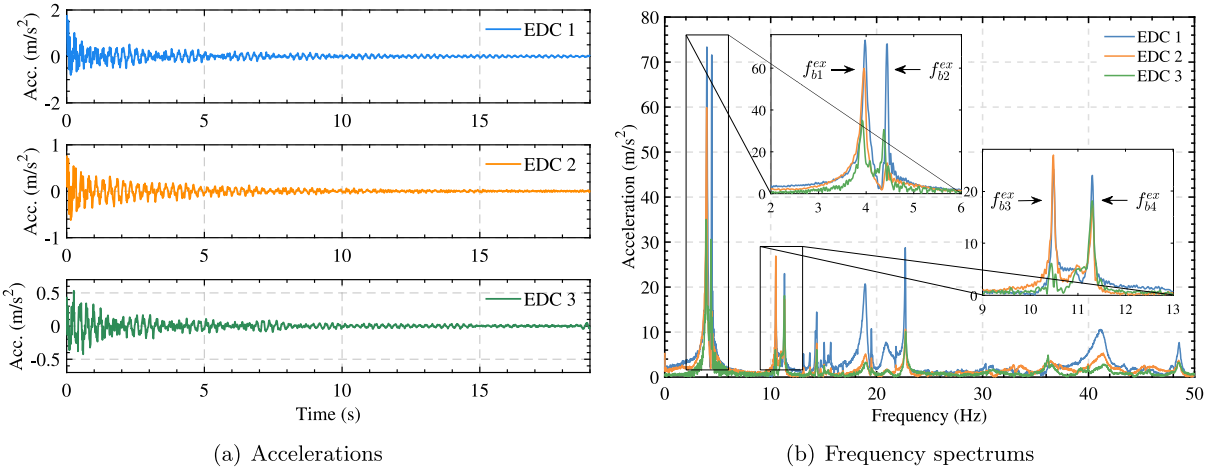


Fig. 14. Footbridge frequency spectrums in different EDCs.

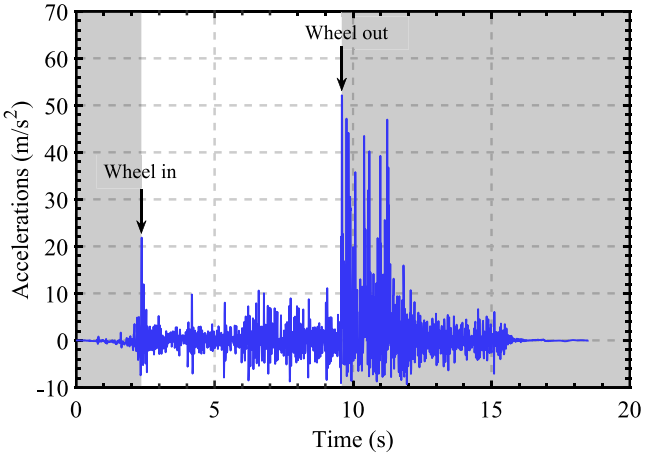


Fig. 15. Data collection.

recorded when the scooter passed the healthy bridge (EDC 0), as shown in Fig. 16. During the forced vibration stage, the scooter traverses the footbridge and vibrates strongly. After the scooter leaves the footbridge, the footbridge vibrates freely because no external forces are applied. The frequency spectra and TFRs of the two vibration stages are shown in Fig. 17.

The forced vibration stage shown in Fig. 17(a) involves the interaction between the scooter and the footbridge, resulting in the footbridge's vibrations being affected by the passing scooter. Despite the introduction of some noisy peaks owing to the presence of the scooter, the direct method (using footbridge vibrations) is still able to effectively identify the footbridge's first several natural frequencies. This highlights the capability of the direct method to capture the dynamic behavior of a footbridge under external loading. From Fig. 17(b), one can obtain the footbridge's natural frequencies when the footbridge is healthy (no additional mass, EDC 0), as listed in Table 2. The following section presents an investigation of the proposed method for detecting footbridge damage using scooter responses.

4.3. Damage detection results and discussions

First, we analyzed scooter acceleration. The accelerations, frequency spectra, and TFRs of the scooter, when the footbridge is in a healthy state (EDC 0), are shown in Fig. 18. A comparison of these results with the simulated results (Fig. 7), we can observe that the accelerations of the scooter during the field tests are more complex and variable. Whether it is the body bounces $\ddot{z}_s^{ex}$, body rotation $\ddot{\theta}_s^{ex}$, or the wheel responses $\ddot{z}_t^{ex}$, it is evident that the frequencies of the footbridge (indicated by dashed yellow lines) cannot be directly identified. Using the STFT with a window length of 100 and an overlapping length of 99, the TRFs of the scooter in the last 5 s were obtained. However, it is clear that the TFRs have been heavily polluted by environmental noises and road roughness, making it impossible to identify any footbridge-related frequency information manually. Therefore, advanced techniques are required to extract critical damage-sensitive features from scooter responses.

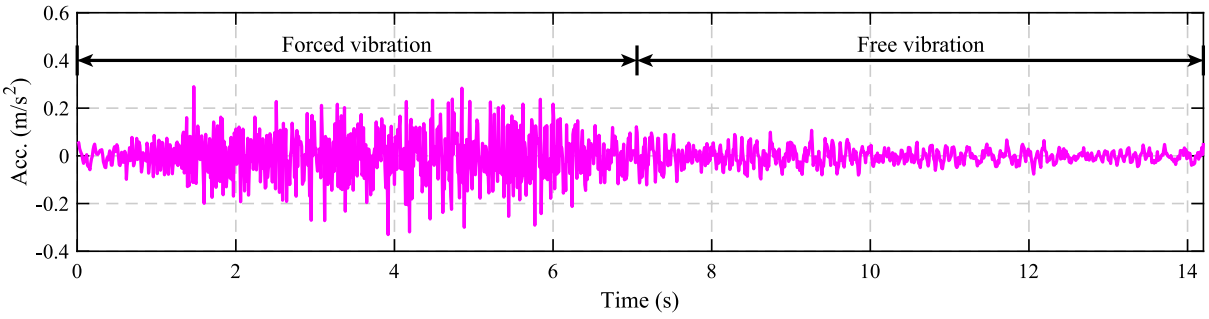


Fig. 16. Vibrations of the footbridge.

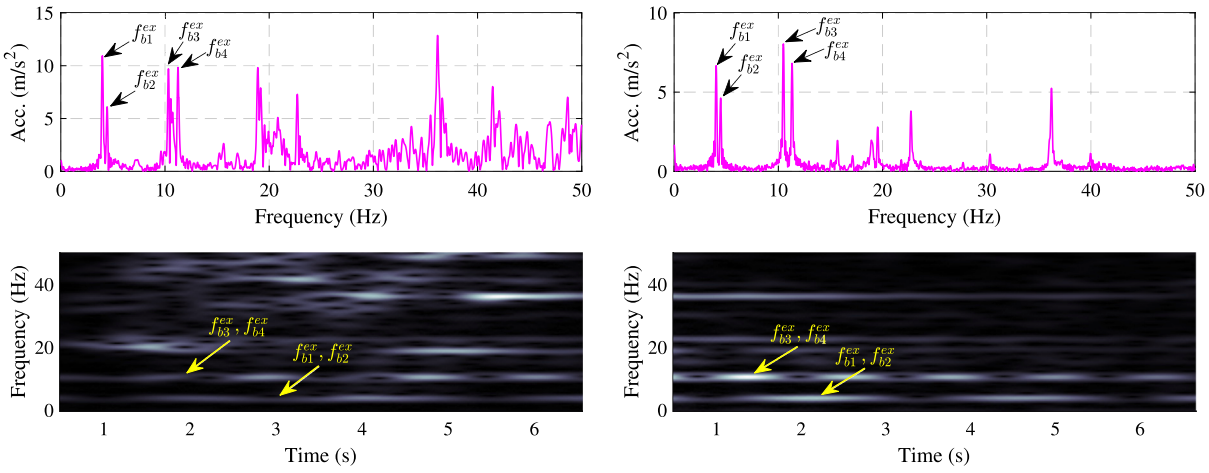


Fig. 17. Footbridge frequency spectrums and TFRs.

In this study, two types of CNNs, as shown in Fig. 1, are employed to detect footbridge damage by analyzing the scooter's responses. To conduct the analysis, we randomly selected 63 runs from EDC 0, resulting in 251 runs in the experimental dataset. For training purposes, 70% of the runs were used, whereas the remaining 30% were used for testing. The CNN configurations used in the simulations were identical. The hyperparameters were selected as follows: batch size, 32; optimizer, Adam; learning rate, $1e^{-6}$; weight decay, $1e^{-5}$; loss function, CE loss; activation function, ReLU; and number of epochs, 400. The CE loss and damage severity prediction accuracy values were calculated using Eq. (8) and are presented in Fig. 19.

The results are shown in Fig. 19(a); one can see that using a 1D CNN significantly decreases training loss after 150 epochs. This suggests that the 1D CNN successfully learned certain features in the scooter response indicative of footbridge damage. Furthermore, the testing loss decreased, although a slight overfitting tendency was observed after 350 epochs. However, both training and testing accuracies continued to improve throughout the training process. Upon reaching 400 epochs, it becomes evident that the testing accuracy reached a value exceeding 70%. In the case of the 2D CNN, as illustrated in Fig. 19(b), both the training and testing losses exhibit earlier declines than those of the 1D CNN. Moreover, the training loss of the 2D CNN approaches a value close to zero. Although there was a slight increase in the testing loss after 150 epochs, the testing accuracy remained consistently higher than 90%. Fig. 20 shows the CMs of the best damage detection results achieved by both types of CNNs for all EDCs.

Upon examination of Fig. 20, it becomes apparent that the 2D CNN, utilizing the scooter's TFRs, outperforms the 1D CNN in terms of accuracy. The 2D CNN demonstrated superior capability to accurately predict footbridge conditions, particularly when the footbridge was damaged. Compared to numerical simulations, which primarily involve uncertainties stemming from Gaussian noise and artificial road roughness, real-world scenarios introduce a higher level of complexity. Consequently, the neural network faces additional challenges distinguishing the damage-sensitive features associated with the footbridge from the scooter's responses. The 2D CNN proved more sensitive to footbridge damage, even considering only the scooter responses within the last 5 s.

To gain a deeper understanding of the superior capabilities of the scooter's TFRs in detecting footbridge damage, it is essential to examine the TFRs of the footbridge. When there is no mass on the footbridge (EDC 0), Fig. 17 clearly shows that the forced vibrations of the footbridge exhibit non-stationary characteristics, particularly evident around 20 Hz and 38 Hz, in comparison to its free vibrations. Even with the introduction of additional masses in EDCs 1–3, the footbridge frequencies decreased, but the aforementioned non-stationary characteristics of the VBI system will remain. Consequently, when sensors are installed on passing scooters, the dynamic information regarding the footbridge in the vibrations of the scooter continues to exhibit time-varying behavior. This offers enhanced opportunities for the 2D CNN to detect damage-sensitive features related to footbridges.

Furthermore, to better interpret the above theories, we utilized Shapley Additive Explanations (SHAP), a commonly used tool for explaining the results of deep learning models. SHAP values are useful for representing the contribution of each feature to the model's predictions [66]. We selected one sample (or run) from each EDC to interpret the factors used by the 2D CNN to determine the bridge health conditions. Fig. 21 presents the SHAP values and probabilities for a TFR image that the 2D CNN predicted as different EDCs. In this figure, the red pixels (with positive SHAP values) increase the output probability

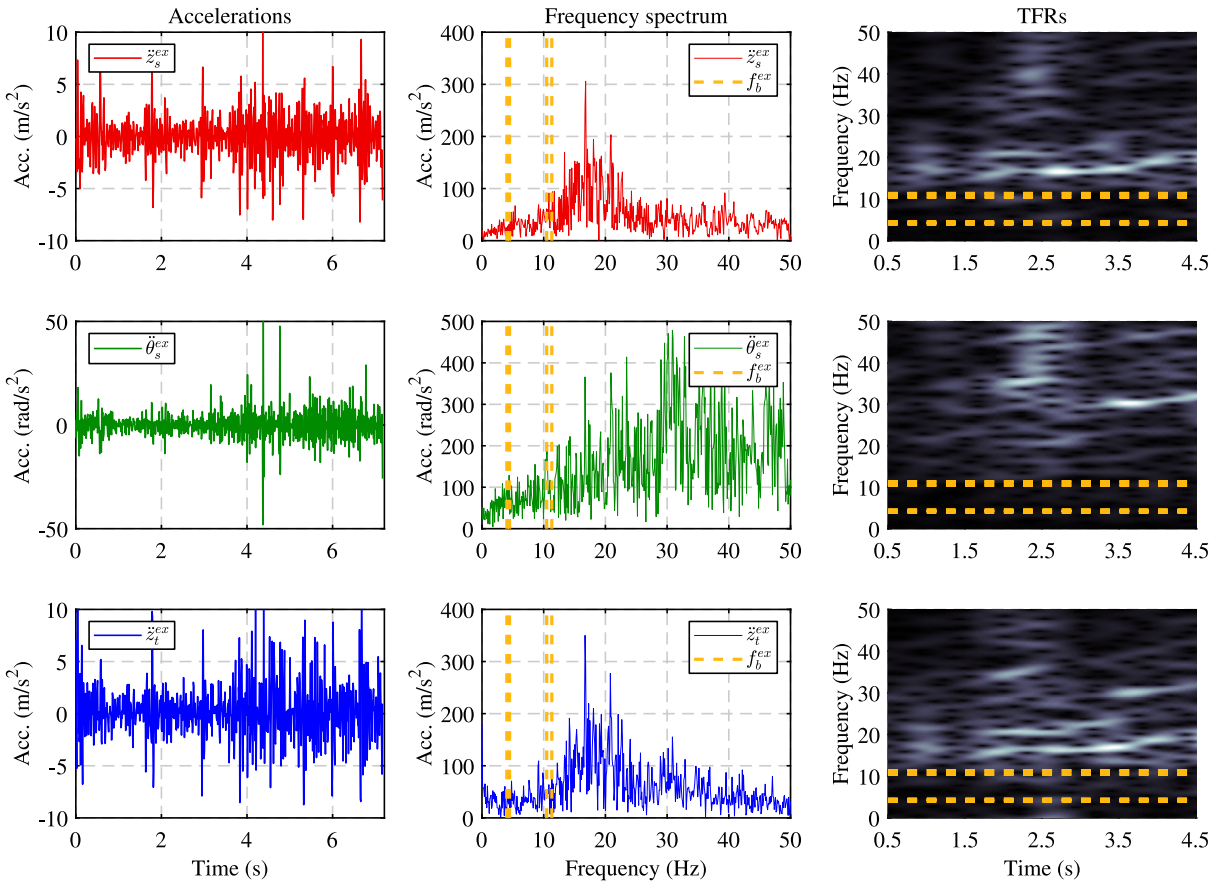
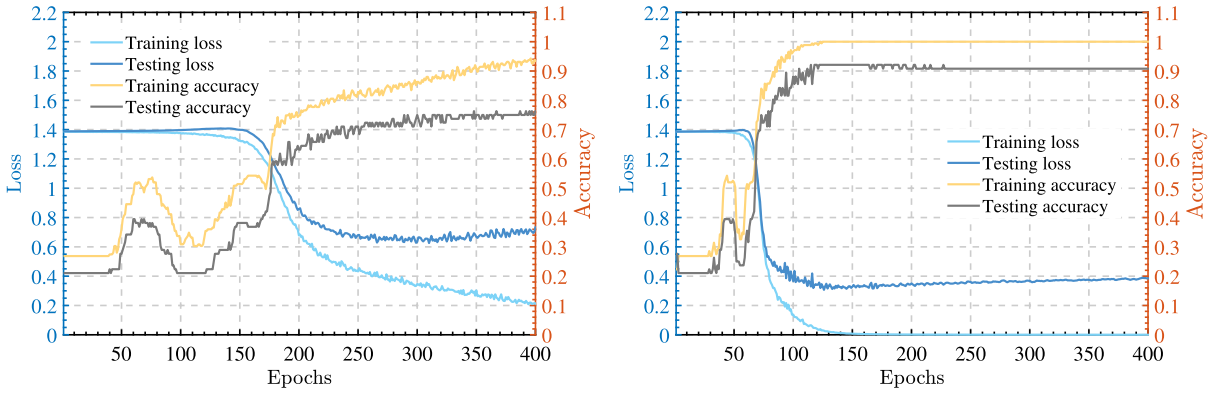


Fig. 18. Responses and TFRs of the scooter in field tests.



(a) Loss and accuracy when 1D CNN is employed

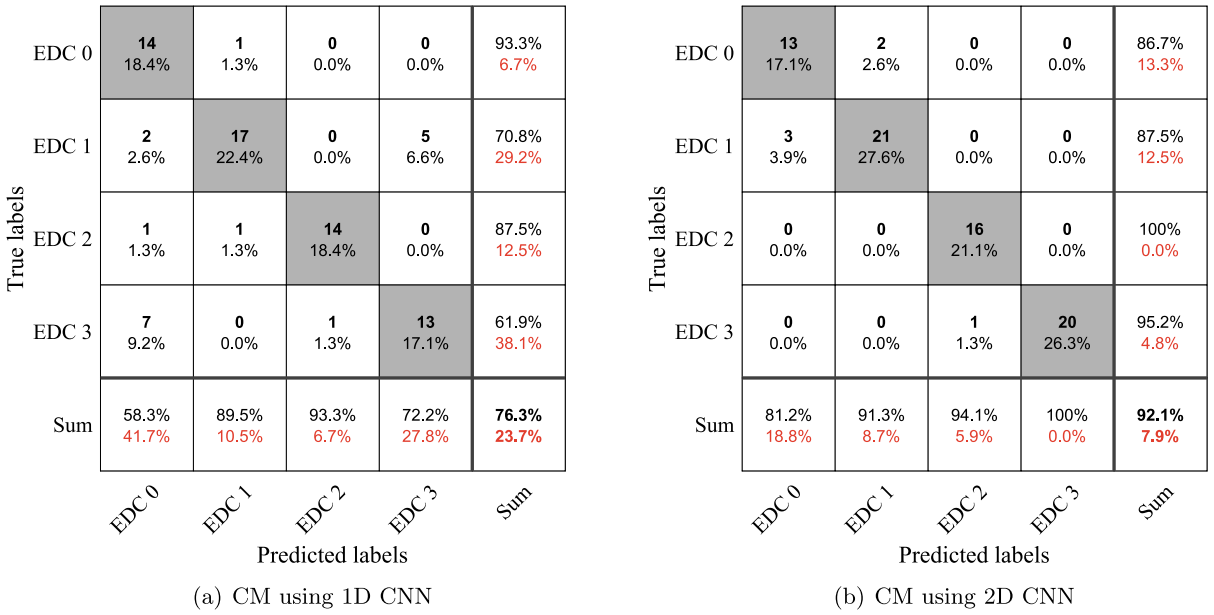
(b) Loss and accuracy when 2D CNN is employed

Fig. 19. Loss and accuracy in field tests using different CNNs.

of the model, whereas the blue pixels (with negative SHAP values) decrease it. Grey pixels have minimal influence on the determination of the CNN model.

From the SHAP value images, it is evident that the majority of significant features, indicated by positive or negative SHAP values, are found within the lower frequency range of 0–30 Hz. This observation is reasonable, considering that the detectable natural frequencies of the footbridge were predominantly below 30 Hz, as shown in Fig. 14(b) when impulse excitations (human jumping) are applied. Specifically, for the sample in EDC 0, image 2 reveals that the significant features are primarily distributed between 0.5–3.5 s, appearing as non-horizontal lines and points. These features correspond to the peaks shown in Fig. 14(b), which are associated with bridge dynamic information. The 2D CNN provided a 99.98% probability of classifying it as EDC 0. Furthermore, it can be observed from images 4 and 5 that certain features, represented by blue pixels, contribute to rejecting it being regarded as EDCs 2 and 3, resulting in extremely low probabilities assigned by the 2D CNN. However, by comparing images 3 and 8, it becomes apparent that the TFRs of EDC 0 also contain vital features that help predict it as EDC 1. Therefore, the 2D CNN assigned a higher probability of 0.01%, although they were far from considering it as EDC 0.

For the sample in EDC 1, image 8 clearly shows that the time-varying frequency responses around 1.5 s are the key identifying features. These features are crucial for distinguishing it from EDC 2, as depicted by the blue pixels in Fig. 9. Due to the similarity of image



(a) CM using 1D CNN (b) CM using 2D CNN

Fig. 20. CMs in field tests using different CNNs.

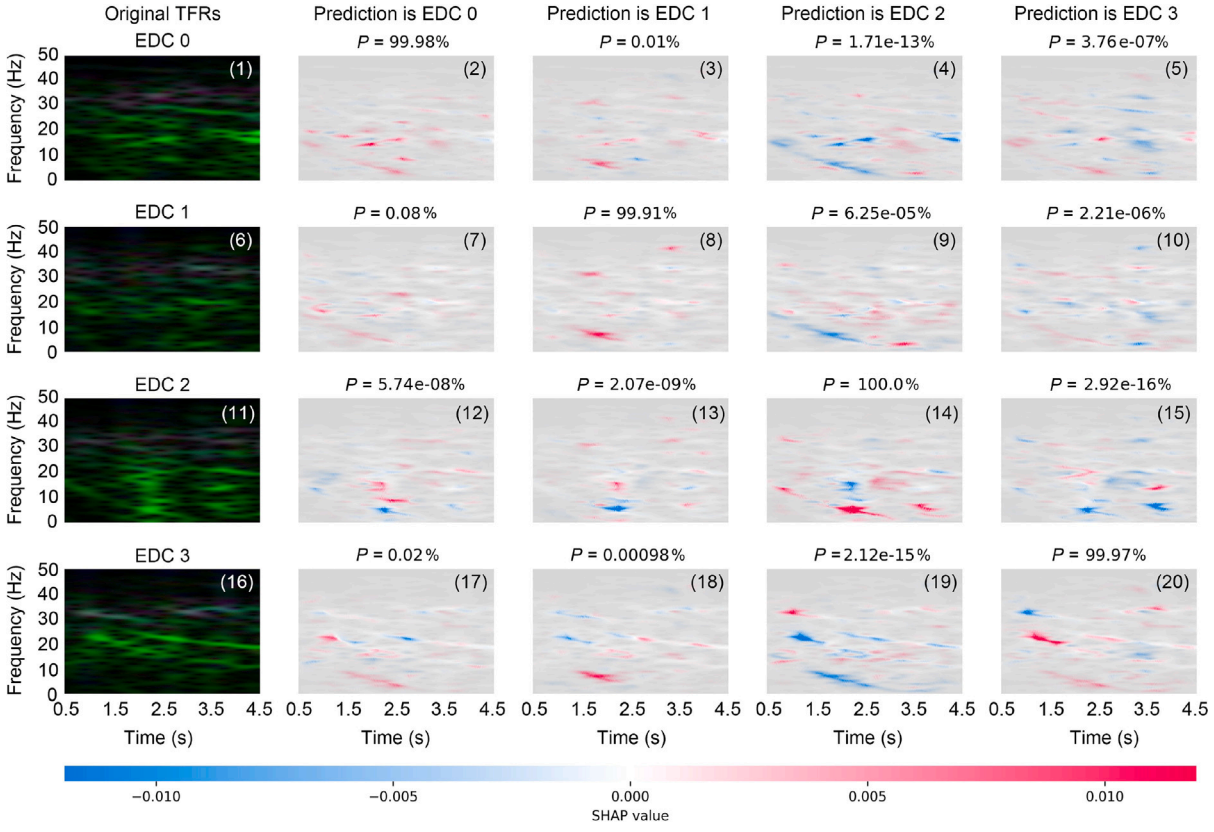


Fig. 21. Interpretation of the 2D CNN using TFRs (*P*: Probability).

8 to image 2 around 1.5 s, the 2D CNN gives a 0.08% probability of regarding it as EDC 0. Additionally, for the sample in EDC 2, determinate features were evident at 2.5 s and 4 s, displaying time-varying characteristics. Turning to the sample in EDC 3, image 20 reveals that the contributing features are located at approximately 4 Hz and 20 Hz. While there is a similar trace in determinate features between 1.5 s and 2.5 s in both EDC 0 (image 2) and EDC 3 (image 20), the higher-frequency time-varying features in image 20 (around 1.5 s) facilitate classifying it as EDC 3 instead of EDC 0. From the above analysis, we can observe that the determinate features utilized by the 2D CNN rely on both time and frequency responses simultaneously. However, when employing the frequency spectra of scooter vibrations, such as in the 1D CNN, the time-varying characteristics of the signals are disregarded, making it challenging to capture crucial damage-sensitive features.

Based on the findings presented above, it can be concluded that for the indirect method when accelerometers are installed on the passing

vehicle, using their TFRs to train the 2D CNN can perform a superior capability in damage detection. This is primarily owing to the non-stationary nature of the response in a VBI system [67–69]; namely, the vehicle and bridge frequencies vary during the interaction process. Similar results regarding the superior capability of using TFRs have also been explored in several studies when sensors are installed on machines or structures when their responses exhibit heavy or slight non-stationary characteristics [44,70,71]. Employing the FFT directly on the vehicle vibrations ignores this time-varying characteristic and is less suitable for damage detection. Therefore, in this study, the 2D CNN was able to effectively capture key damage-sensitive features in the vibrations of the scooter, resulting in a better performance than the 1D CNN.

5. Conclusions and future work

This paper presents a framework for detecting and classifying footbridge damage by analyzing micromobility responses equipped with smartphones and deep learning techniques. The TFRs of the scooter's vibrations were innovatively utilized by the 2D CNN to predict the footbridge's damage severity in comparison with the 1D CNN using the scooter's frequency spectra. The effectiveness of the proposed framework was evaluated using a numerical model of scooter-footbridge interactions and real-world field tests. Several concluding remarks are presented below.

(1) In numerical simulations, both the 1D and 2D CNNs demonstrate a notable capability to detect and classify footbridge damage severities using vibrations from passing scooters. The accuracy achieved by the two CNNs is relatively high (91.7% and 93.5%, respectively) even when the scooter's responses are affected by artificial road roughness and Gaussian noises.
(2) In field tests with stochastic influence factors, the superior capability of the 2D CNN is evident compared to the 1D CNN. The accuracy of damage classification decreases to 76.3% for the 1D CNN, whereas the 2D CNN maintains a high accuracy of 92.1% by utilizing the TFRs of the scooter's vibrations.
(3) The superior capability of the 2D CNN using TFRs can be attributed to the non-stationary nature of the VBI system responses. Further investigations using SHAP values reveal that the damage-sensitive features of the footbridge vary in the scooter's vibrations over time.

Although some promising findings are obtained, there are certain limitations in the current research. One is the use of a smartphone attached to the front wheel of a scooter, which is not commonly employed in daily life and may impede its use in general engineering applications. Additionally, this study did not consider factors such as the effects of temperature, significant changes in road roughness, and the presence of pedestrians despite their significance in long-term footbridge monitoring. Thirdly, the proposed approach is supervised, and the scarcity of damaged scenarios can limit its applications in engineering. In our future studies, we will investigate the utilization of generative and digital twin models to generate various types of damage data for model training, as well as unsupervised learning without any labeled data. Moreover, we aim to address limitations by investigating the effectiveness of using smartphones on standing slabs and handlebars and exploring the aforementioned influential factors. Owing to easy access to visual inspection, it is also worthwhile conducting comprehensive investigations into the combination of hybrid proposed and vision-based methods for inspecting short-span footbridges.

CRediT authorship contribution statement

Zhenkun Li: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Conceptualization. **Yifu Lan:** Writing – review & editing. **Weiwei Lin:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research is financially sponsored by the Jane and Aatos Erkko Foundation in Finland (Decision number: 210018) and Aalto University, Finland (research project funding in ENG 2022). The help from colleagues on the field tests is highly appreciated.

References

[1] P. Negi, R. Kromanis, A.G. Dorée, K.M. Wijnberg, Structural health monitoring of inland navigation structures and ports: a review on developments and challenges, Struct. Health Monit. 23 (1) (2023) 605–645, http://dx.doi.org/10.1177/14759217231170742.
[2] Z. Peng, J. Li, H. Hao, Computer vision-based displacement identification and its application to bridge condition assessment under operational conditions, Smart Constr. 1 (1) (2024) 0003, http://dx.doi.org/10.55092/sc20240003.
[3] R. Hou, Y. Xia, Review on the new development of vibration-based damage identification for civil engineering structures: 2010–2019, J. Sound Vib. 491 (2021) 115741, http://dx.doi.org/10.1016/j.jsv.2020.115741.
[4] C. Zhang, A.A. Mousavi, S.F. Masri, G. Gholipour, K. Yan, X. Li, Vibration feature extraction using signal processing techniques for structural health monitoring: a review, Mech. Syst. Signal Process. 177 (2022) 109175, http://dx.doi.org/10.1016/j.ymssp.2022.109175.
[5] K. Feng, A. González, M. Casero, A kNN algorithm for locating and quantifying stiffness loss in a bridge from the forced vibration due to a truck crossing at low speed, Mech. Syst. Signal Process. 154 (2021) 107599, http://dx.doi.org/10.1016/j.ymssp.2020.107599.
[6] J. Shu, C. Zhang, Y. Gao, Y. Niu, A multi-task learning-based automatic blind identification procedure for operational modal analysis, Mech. Syst. Signal Process. 187 (2023) 109959, http://dx.doi.org/10.1016/j.ymssp.2022.109959.
[7] Y.B. Yang, C.W. Lin, J.D. Yau, Extracting bridge frequencies from the dynamic response of a passing vehicle, in: H. Takemiya (Ed.), J. Sound Vib. 272 (3–5) (2004) 471–493, http://dx.doi.org/10.1016/S0022-460X(03)00378-X.
[8] C.W. Lin, Y.B. Yang, Use of a passing vehicle to scan the fundamental bridge frequencies: an experimental verification, Eng. Struct. 27 (13) (2005) 1865–1878, http://dx.doi.org/10.1016/j.engstruct.2005.06.016.
[9] A. Malekjafarian, R. Corbally, W. Gong, A review of mobile sensing of bridges using moving vehicles: progress to date, challenges and future trends, Structures 44 (2022) 1466–1489, http://dx.doi.org/10.1016/j.istruc.2022.08.075.
[10] Z.L. Wang, J.P. Yang, K. Shi, H. Xu, F.Q. Qiu, Y.B. Yang, Recent advances in researches on vehicle scanning method for bridges, Int. J. Struct. Stab. Dyn. 22 (15) (2022) 2230005, http://dx.doi.org/10.1142/S0219455422300051.
[11] Y.B. Yang, Z. Li, Z.L. Wang, K. Shi, H. Xu, F.Q. Qiu, J.F. Zhu, A novel frequency-free movable test vehicle for retrieving modal parameters of bridges: theory and experiment, Mech. Syst. Signal Process. 170 (2022) 108854, http://dx.doi.org/10.1016/j.ymssp.2022.108854.
[12] J.P. Yang, W.C. Lee, Damping effect of a passing vehicle for indirectly measuring bridge frequencies by EMD technique, Int. J. Struct. Stab. Dyn. 18 (01) (2017) 1850008, http://dx.doi.org/10.1142/S0219455418500086.
[13] Y.B. Yang, J. Chen, H. Xu, Normalized formula for removing damping effect in recovering bridge mode shapes using a moving and a stationary vehicle, J. Sound Vib. 573 (2024) 118219, http://dx.doi.org/10.1016/j.jsv.2023.118219.
[14] Z. Li, Y. Lan, W. Lin, Indirect damage detection for bridges using sensing and temporarily parked vehicles, Eng. Struct. 291 (2023) 116459, http://dx.doi.org/10.1016/j.engstruct.2023.116459.
[15] R. Corbally, A. Malekjafarian, Detecting changes in the structural behaviour of a laboratory bridge model using the contact-point response of a passing vehicle, J. Struct. Integr. Maint. 8 (4) (2023) 226–238, http://dx.doi.org/10.1080/24705314.2023.2230399.
[16] Y. He, J.P. Yang, Z. Yan, Enhanced identification of bridge modal parameters using contact residuals from three-connected vehicles: theoretical study, Structures 54 (2023) 1320–1335, http://dx.doi.org/10.1016/j.istruc.2023.05.112.
[17] X. Jian, Y. Xia, L. Sun, An indirect method for bridge mode shapes identification based on wavelet analysis, Struct. Control Health Monit. 27 (12) (2020) e2630, http://dx.doi.org/10.1002/stc.2630.

[18] J. Zhang, T.H. Yi, C.X. Qu, Q. Han, Y.F. Wang, X.D. Mei, Experimental studies of extracting bridge mode shapes by response of a moving vehicle, J. Bridge Eng. 28 (11) (2023) 04023076, http://dx.doi.org/10.1061/JBENF2.BEENG-6243.
[19] Y. Zhang, L. Wang, Z. Xiang, Damage detection by mode shape squares extracted from a passing vehicle, J. Sound Vib. 331 (2) (2012) 291–307, http://dx.doi.org/10.1016/j.jsv.2011.09.004.
[20] H. Xu, M. Yang, J.P. Yang, Z.L. Wang, K. Shi, Y.B. Yang, Vehicle scanning method for bridges enhanced by dual amplifiers, Struct. Control Health Monit. 2023 (2023) 6906855, http://dx.doi.org/10.1155/2023/6906855.
[21] Y. He, J.P. Yang, Using acceleration residual spectrum from single two-axle vehicle at contact points to extract bridge frequencies, Eng. Struct. 266 (2022) 114538, http://dx.doi.org/10.1016/j.engstruct.2022.114538.
[22] C.W. Kim, K.C. Chang, S. Inoue, S. Hasegawa, P.J. McGetrick, Investigating feasibility of drive-by bridge monitoring by laboratory experiments, in: Mechanics of Structures and Materials: Advancements and Challenges - Proceedings of the 24th Australasian Conference on the Mechanics of Structures and Materials (ACMSM24 2016), Perth, Australia, 6-9 December, 2016, pp. 1477–1484, http://dx.doi.org/10.1201/9781315226460.
[23] M.Y. Cheng, K.W. Liao, Y.F. Chiu, Y.W. Wu, S.H. Yeh, T.C. Lin, Automated mobile vibration measurement and signal analysis for bridge scour prevention and warning, Autom. Constr. 134 (2022) 104063, http://dx.doi.org/10.1016/j.autcon.2021.104063.
[24] W. Locke, J. Sybrandt, L. Redmond, I. Safro, S. Atamturktur, Using drive-by health monitoring to detect bridge damage considering environmental and operational effects, J. Sound Vib. 468 (2020) 115088, http://dx.doi.org/10.1016/j.jsv.2019.115088.
[25] F. Cerda, S. Chen, J. Bielak, J.H. Garrett, P. Rizzo, J. Kovacevic, Indirect structural health monitoring of a simplified laboratory-scale bridge model, Smart Struct. Syst. 13 (5) (2014) 849–868, http://dx.doi.org/10.12989/sss.2014.13.5.849.
[26] Z. Li, W. Lin, Y. Zhang, Drive-by bridge damage detection using Mel-frequency cepstral coefficients and support vector machine, Struct. Health Monit. 22 (5) (2023) 3302–3319, http://dx.doi.org/10.1177/14759217221150932.
[27] J. Liu, S. Chen, M. Bergés, J. Bielak, J.H. Garrett, J. Kovačević, H.Y. Noh, Diagnosis algorithms for indirect structural health monitoring of a bridge model via dimensionality reduction, Mech. Syst. Signal Process. 136 (2020) 106454, http://dx.doi.org/10.1016/j.ymssp.2019.106454.
[28] P. Cheema, M.M. Alamdari, K.C. Chang, C.W. Kim, M. Sugiyama, A drive-by bridge inspection framework using non-parametric clusters over projected data manifolds, Mech. Syst. Signal Process. 180 (2022) 109401, http://dx.doi.org/10.1016/j.ymssp.2022.109401.
[29] Q. Mei, M. Gül, M. Boay, Indirect health monitoring of bridges using Mel-frequency cepstral coefficients and principal component analysis, Mech. Syst. Signal Process. 119 (2019) 523–546, http://dx.doi.org/10.1016/j.ymssp.2018.10.006.
[30] Y. Lan, Y. Zhang, W. Lin, Diagnosis algorithms for indirect bridge health monitoring via an optimized AdaBoost-linear SVM, Eng. Struct. 275 (2023) 115239, http://dx.doi.org/10.1016/j.engstruct.2022.115239.
[31] S. Mokalled, W. Locke, O. Abuodeh, L. Redmond, C. McMahan, Drive-by health monitoring of highway bridges using Bayesian estimation technique for damage classification, Struct. Control Health Monit. 29 (6) (2022) e2944, http://dx.doi.org/10.1002/stc.2944.
[32] A. Malekjafarian, F. Golpayegani, C. Moloney, S. Clarke, A machine learning approach to bridge-damage detection using responses measured on a passing vehicle, Sensors 19 (18) (2019) 4035, http://dx.doi.org/10.3390/s19184035.
[33] R. Corbally, A. Malekjafarian, A data-driven approach for drive-by damage detection in bridges considering the influence of temperature change, Eng. Struct. 253 (2022) 113783, http://dx.doi.org/10.1016/j.engstruct.2021.113783.
[34] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90, http://dx.doi.org/10.1145/3065386.
[35] R. Fu, M. Cao, D. Novák, X. Qian, N.F. Alkayem, Extended efficient convolutional neural network for concrete crack detection with illustrated merits, Autom. Constr. 156 (2023) 105098, http://dx.doi.org/10.1016/j.autcon.2023.105098.
[36] G. Liu, Y. Niu, W. Zhao, Y. Duan, J. Shu, Data anomaly detection for structural health monitoring using a combination network of GANomaly and CNN, Smart Struct. Syst. 29 (1) (2022) 53–62, http://dx.doi.org/10.12989/sss.2022.29.1.053.
[37] W. Ding, H. Yang, K. Yu, J. Shu, Crack detection and quantification for concrete structures using UAV and transformer, Autom. Constr. 152 (2023) 104929, http://dx.doi.org/10.1016/j.autcon.2023.104929.
[38] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, D.J. Inman, Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks, J. Sound Vib. 388 (2017) 154–170, http://dx.doi.org/10.1016/j.jsv.2016.10.043.
[39] J. Shu, C. Zhang, X. Chen, Y. Niu, Model-informed deep learning strategy with vision measurement for damage identification of truss structures, Mech. Syst. Signal Process. 196 (2023) 110327, http://dx.doi.org/10.1016/j.ymssp.2023.110327.
[40] D. Hajializadeh, Deep learning-based indirect bridge damage identification system, Struct. Health Monit. 22 (2) (2022) 897–912, http://dx.doi.org/10.1177/14759217221087147.
[41] R. Corbally, A. Malekjafarian, A deep-learning framework for classifying the type, location, and severity of bridge damage using drive-by measurements, Comput.-Aided Civ. Infrastruct. Eng. 39 (6) (2024) 852–871, http://dx.doi.org/10.1111/mice.13104.
[42] W.H. Hu, E. Caetano, á. Cunha, Structural health monitoring of a stress-ribbon footbridge, Eng. Struct. 57 (2013) 578–593, http://dx.doi.org/10.1016/j.engstruct.2012.06.051.
[43] I. Belykh, M. Bocian, A.R. Champneys, K. Daley, R. Jeter, J.H.G. Macdonald, A. McRobie, Emergence of the London millennium bridge instability without synchronisation, Nature Commun. 12 (1) (2021) 7223, http://dx.doi.org/10.1038/s41467-021-27568-y.
[44] M. Jamshidi, M. El-Badry, Structural damage severity classification from time-frequency acceleration data using convolutional neural networks, Structures 54 (2023) 236–253, http://dx.doi.org/10.1016/j.istruc.2023.05.009.
[45] Z. Peng, J. Li, H. Hao, Development and experimental verification of an IoT sensing system for drive-by bridge health monitoring, Eng. Struct. 293 (2023) 116705, http://dx.doi.org/10.1016/j.engstruct.2023.116705.
[46] Q. Mei, M. Gül, N. Shirzad-Ghaleroudkhani, Towards smart cities: crowdsensing-based monitoring of transportation infrastructure using in-traffic vehicles, J. Civ. Struct. Health Monit. 10 (4) (2020) 653–665, http://dx.doi.org/10.1007/s13349-020-00411-6.
[47] S. Quqa, P.F. Giordano, M.P. Limongelli, Shared micromobility-driven modal identification of urban bridges, Autom. Constr. 134 (2022) 104048, http://dx.doi.org/10.1016/j.autcon.2021.104048.
[48] B.Y. Zhang, Y.Q. Ni, A data-driven sensor placement strategy for reconstruction of mode shapes by using recurrent Gaussian process regression, Eng. Struct. 284 (2023) 115998, http://dx.doi.org/10.1016/j.engstruct.2023.115998.
[49] J. Hou, Z. Li, Ł. Jankowski, S. Wang, Estimation of virtual masses for structural damage identification, Struct. Control Health Monit. 27 (8) (2020) e2585, http://dx.doi.org/10.1002/stc.2585.
[50] H. Xu, Y.H. Liu, M. Yang, D.S. Yang, Y.B. Yang, Mode shape construction for bridges from contact responses of a two-axle test vehicle by wavelet transform, Mech. Syst. Signal Process. 195 (2023) 110304, http://dx.doi.org/10.1016/j.ymssp.2023.110304.
[51] Z. Li, W. Lin, Y. Zhang, Bridge frequency scanning using the contact-point response of an instrumented 3D vehicle: theory and numerical simulation, Struct. Control Health Monit. 2023 (2023) 3924349, http://dx.doi.org/10.1155/2023/3924349.
[52] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2015, http://dx.doi.org/10.48550/arXiv.1409.1556, ArXiv Preprint.
[53] G. Kolappan Geetha, S.H. Sim, Fast identification of concrete cracks using 1D deep learning and explainable artificial intelligence-based analysis, Autom. Constr. 143 (2022) 104572, http://dx.doi.org/10.1016/j.autcon.2022.104572.
[54] O. Abdeljaber, O. Avci, M.S. Kiranyaz, B. Boashash, H. Sodano, D.J. Inman, 1-D CNNs for structural damage detection: verification on a structural health monitoring benchmark data, Neurocomputing 275 (2018) 1308–1317, http://dx.doi.org/10.1016/j.neucom.2017.09.069.
[55] Y. Matsumoto, M.J. Griffin, Mathematical models for the apparent masses of standing subjects exposed to vertical whole-body vibration, J. Sound Vib. 260 (3) (2003) 431–451, http://dx.doi.org/10.1016/S0022-460X(02)00941-0.
[56] Z. Li, Y. Lan, W. Lin, Indirect frequency identification of footbridges with pedestrians using the contact-point response of shared scooters, J. Bridge Eng. 29 (6) (2024) 04024036, http://dx.doi.org/10.1061/JBENF2.BEENG-6344.
[57] H. Xu, Y.H. Liu, J. Chen, D.S. Yang, Y.B. Yang, Novel formula for determining bridge damping ratio from two wheels of a scanning vehicle by wavelet transform, Mech. Syst. Signal Process. 208 (2024) 111026, http://dx.doi.org/10.1016/j.ymssp.2023.111026.
[58] Mechanical vibration — road surface profiles — reporting of measured data, 2023, https://www.iso.org/standard/71202.html. (Accessed 16 November 2023).
[59] M.Z. Sarwar, D. Cantero, Deep autoencoder architecture for bridge damage assessment using responses from several vehicles, Eng. Struct. 246 (2021) 113064, http://dx.doi.org/10.1016/j.engstruct.2021.113064.
[60] O. Markogiannaki, A. Arailopoulos, D. Giagopoulos, C. Papadimitriou, Vibration-based damage localization and quantification framework of large-scale truss structures, Struct. Health Monit. 22 (2) (2022) 1376–1398, http://dx.doi.org/10.1177/14759217221100443.
[61] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, PyTorch: an imperative style, high-performance deep learning library, 2019, http://dx.doi.org/10.48550/arXiv.1912.01703, ArXiv Preprint.
[62] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, 2017, http://dx.doi.org/10.48550/arXiv.1412.6980, ArXiv Preprint.
[63] X. Ye, Y. Cao, A. Liu, X. Wang, Y. Zhao, N. Hu, Parallel convolutional neural network toward high efficiency and robust structural damage identification, Struct. Health Monit. 22 (6) (2023) 3805–3826, http://dx.doi.org/10.1177/14759217231158786.

[64] Q. Mei, M. Gül, N. Shirzad-Ghaleroudkhani, Towards smart cities: crowdsensing-based monitoring of transportation infrastructure using in-traffic vehicles, J. Civ. Struct. Health Monit. 10 (4) (2020) 653–665, http://dx.doi.org/10.1007/s13349-020-00411-6.

[65] Z. Li, Y. Lan, W. Lin, Investigation of frequency-domain dimension reduction for A2M-based bridge damage detection using accelerations of moving vehicles, Materials 16 (5) (2023) 1872, http://dx.doi.org/10.3390/ma16051872.

[66] Welcome to the SHAP documentation, 2023, https://shap.readthedocs.io/en/latest. (Accessed 16 November 2023).

[67] Y.B. Yang, M.C. Cheng, K.C. Chang, Frequency variation in vehicle–bridge interaction systems, Int. J. Struct. Stab. Dyn. 13 (02) (2013) 1350019, http://dx.doi.org/10.1142/S0219455413500193.

[68] J. Li, X. Zhu, S.S. Law, B. Samali, Time-varying characteristics of bridges under the passage of vehicles using synchroextracting transform, Mech. Syst. Signal Process. 140 (2020) 106727, http://dx.doi.org/10.1016/j.ymssp.2020.106727.

[69] C. Tan, H. Zhao, E.J. OBrien, N. Uddin, C.W. Kim, Exploring time-varying characteristics in drive-by bridge frequency extraction with the second-order synchrosqueezing transform, J. Bridge Eng. 28 (4) (2023) 04023010, http://dx.doi.org/10.1061/JBENF2.BEENG-5979.

[70] Y. Zhang, K. Xing, R. Bai, D. Sun, Z. Meng, An enhanced convolutional neural network for bearing fault diagnosis based on time–frequency image, Measurement 157 (2020) 107667, http://dx.doi.org/10.1016/j.measurement.2020.107667.

[71] Z. Chen, Y. Wang, J. Wu, C. Deng, K. Hu, Sensor data-driven structural damage detection based on deep convolutional neural networks and continuous wavelet transform, Appl. Intell. 51 (8) (2021) 5598–5609, http://dx.doi.org/10.1007/s10489-020-02092-6.