

Data-Driven Structural Health Monitoring Using Feature Fusion and Hybrid Deep Learning

Hung V. Dang[✉], Hoa Tran-Ngoc[✉], Tung V. Nguyen, T. Bui-Tien[✉], Guido De Roeck, and Huan X. Nguyen[✉], *Senior Member, IEEE*

Abstract—Smart structural health monitoring (SHM) for large-scale infrastructure is an intriguing subject for engineering communities thanks to its significant advantages such as timely damage detection, optimal maintenance strategy, and reduced resource requirement. Yet, it is a challenging topic as it requires handling a large amount of collected sensors data continuously, which is inevitably contaminated by random noises. Therefore, this study developed a practical end-to-end framework that makes use of physical features embedded in raw data and an elaborated hybrid deep learning model, namely 1-DCNN-LSTM, featuring two algorithms—convolutional neural network (CNN) and long-short term memory (LSTM). In order to extract relevant features from sensory data, the method combines various signal processing techniques such as the autoregressive model, discrete wavelet transform, and empirical mode decomposition. The hybrid deep learning 1-DCNN-LSTM is designed based on the CNN’s capacity of capturing local information and the LSTM network’s prominent ability to learn long-term dependencies. Through three case studies involving both experimental and synthetic data sets, it is demonstrated that the proposed approach achieves highly accurate damage detection, as accurate as the powerful 2-D CNN, but with a lower time and memory complexity, making it suitable for real-time SHM.

Manuscript received July 22, 2020; revised August 28, 2020; accepted October 24, 2020. Date of publication November 13, 2020; date of current version October 6, 2021. This article was recommended for publication by Associate Editor C.-Y. Lee and Editor F.-T. Cheng upon evaluation of the reviewers’ comments. This work was supported in part by an Institutional Links Grant, ID 429715093, under the Newton Program Vietnam partnership and in part by the U.K. Department of Business, Energy and Industrial Strategy (BEIS) and delivered by the British Council. For further information, please refer to: www.newtonfund.ac.uk. (*Corresponding author: Hung V. Dang*)

Hung V. Dang is with the London Digital Twin Research Centre, Faculty of Science and Technology, Middlesex University, NW4 4BT London, U.K., and also with the Faculty of Building and Industrial Construction, National University of Civil Engineering, Hanoi, Vietnam (e-mail: D.viethung@mdx.ac.uk).

Hoa Tran-Ngoc is with the Department of Bridge and Tunnel Engineering, Faculty of Civil Engineering, University of Transport and Communications, Hanoi, Vietnam, and also with the Department of Electrical Energy, Metals, Mechanical Constructions, and Systems, Faculty of Engineering and Architecture, Ghent University, 9000 Ghent, Belgium.

Tung V. Nguyen is with the Modeling Simulation Team, Schlumberger, 92140 Clamart, France.

T. Bui-Tien is with the Department of Bridge and Tunnel Engineering, Faculty of Civil Engineering, University of Transport and Communications, Hanoi, Vietnam.

Guido De Roeck is with the Department of Civil Engineering, KU Leuven, B-3001 Leuven, Belgium.

Huan X. Nguyen is with the London Digital Twin Research Center, Faculty of Science and Technology, Middlesex University, NW4 4BT London, U.K. (e-mail: H.Nguyen@mdx.ac.uk).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TASE.2020.3034401>.

Digital Object Identifier 10.1109/TASE.2020.3034401

Note to Practitioners—This article aims to develop a practical data-driven method for automatically monitoring the operational state of structures. In order to achieve consistently and highly accurate results in performing different tasks for diverse structures, we combine underlying features in both time and frequency domains extracted from measured signal vibration data. Three popular data featuring methods are combined to achieve the diversity gain which would not be possible with each individual method. As the vibration is usually measured by long time-series signals, the most efficient deep learning architecture for time-series signal, namely long-short term memory (LSTM), is considered for this work. Besides, each structure has its own dynamic properties, i.e., eigenfrequencies, around which the most relevant information is in the frequency domain, thus convolutional neural network specifically designed for capturing local information is used in combination with LSTM, forming a hybrid deep learning architecture. The applicability and effectiveness of the proposed approach are supported by three case studies with different types of structures, showing highly accurate damage detection with reduced resource requirements. These advantages can be valuable for developing a model for live monitoring of structural health in the future life-line infrastructures.

Index Terms—Damage detection, deep learning (DL), dynamic analysis, signal processing, structural health monitoring (SHM), vibration.

I. INTRODUCTION

LARGE-SCALE civil structures are expensive assets playing a vital role in society [1] as they ensure smooth transportation and improve the quality of people’s daily life. However, they are permanently exposed to various unpredicted excitations involving wind loads, vehicular loads, accidental loads, environmental changes, and even earthquakes. Toward intelligent and real-time monitoring, the data-driven model recently emerges as a promising alternative to other techniques, which provides unprecedented advantages such as timely detecting damage, predicting structural behaviors in extreme scenarios, and optimizing maintenance strategies. However, it is challenging to design a high-performance algorithm dealing with a large volume of data for long-term structural health monitoring (SHM).

The autoregressive (AR) model belonging to the time-domain class is widely used when working with sensor data in many sectors such as civil, mechanical engineering, information technology, healthcare, etc. Sohn *et al.* [2] used AR in combination with the X-bar control chart to handle vibration-based damage detection problems of concrete bridge columns. It is shown that the method is able to identify all investigated damage levels in an unsupervised learning

fashion, which is significantly useful in real application because of the scarcity of data related to different damage scenarios. Entezami *et al.* [3] developed a fast unsupervised approach for SHM having the capability of dealing with large vibration data based on the AR models for feature extraction and Kullback–Leibler distance for classification. The applicability of the method was supported through both numerical simulation and experimental data sets, with emphasis on its computational efficiency and highly accurate damage localization. Carden and Brownjohn [4] developed a derived AR technique called autoregressive moving average (ARMA) model, to assess respective conditions of a structure in service, a.k.a. the level 1 of SHM. The method was proved to distinguish the health state from various damaged ones through three data sets acquired from the laboratory IASC–ASCE benchmark structure and two real-world structures, i.e., the Z24 bridge and the Singapore–Malaysia Second Link Bridge. However, the AR-based methods require mathematically elaborated criteria and manual thresholds to classify time-series into corresponding classes. These thresholds are usually sensitive not only to structures' properties but also to characteristics of excitation and environmental parameters.

In reality, the correctness of a time-domain technique may be impaired by unwanted noise, then techniques in time–frequency domain are developed, of which the wavelet transform (WT) is commonly appreciated by its ability to provide a proper representation of data with multiple time and frequency resolutions. Pnevmatikos *et al.* [5] achieved high accurate results in damage localization of a steel frame structure by estimating the difference of wavelet coefficients extracted from acceleration measurements of undamaged and damaged structures subjected to an earthquake. For long term SHM of bridges, Zeng *et al.* [6] employed WT to discern characteristics in both time and frequency domain of structures responses under external excitation such as temperature variation and traffic loads. Recently, Ma *et al.* [7] combined WT and independent component analysis to identify sudden stiffness loss of a nonlinear hysteretic structure in a fast and accurate fashion.

On the other hand, an adaptive technique called empirical mode decomposition (EMD), which is able to retain local characteristics of data, has received increasing interest. For bridge damage detection under vehicular load, Obrien *et al.* [8] found that intrinsic mode functions (IMFs) corresponding to pseudo-frequency, obtained from EMD are sensitive to damage and can serve as effective indicators for damage localization. Sadhu [9] investigated a hybrid EMD method to address the potential mode mixing issue of the traditional EMD method. The proposed method was then applied to perform modal identification and achieved accurate results, for example, modal assurance criteria of higher than 0.98, despite measurement noise and closely spaced modes. In [10], Xiao *et al.*, developed an extension of the EMD to analyze the acceleration response of the Chulitna River bridge in Alaska to a moving vehicle, which could provide clear transient frequencies with high resolution without the need of linearity assumptions required by Fourier transform-based methods.

Another way to work with measured time series data is to convert them into 2-D graphical representations, then using computer vision algorithms to detect the corresponding states of the physical object. Tang *et al.* [11] applied a 2-D convolutional neural network (2-DCNN) to the anomaly detection problem of a long-span cable-stayed bridge, utilizing 2-D representation of accelerometer sensor signals. Their results suggested that the 2-DCNN approach performed better than the existing method in structural anomaly detection, and was scalable to include more signal data from multiple measurements. This idea is also supported by research in other domains, for example, the works of [12] and [13] in machine fault diagnosis.

Although 2-DCNN is a powerful algorithm for structural damage detection (SDD) tasks, the network requires 2-D time–frequency representation of the 1-D sensor time-series, and subsequent image management process including proper selection of parameters for image conversion, labeling images, splitting images into separate training/validation/testing folder, moreover when performing a new task with same data or new label incorporated, the whole process related to image needs to repeat, or even to duplicate existing images. Besides, reviewed signal processing techniques have their own advantages but also drawbacks such as manual thresholds, high time complexity, and less flexibility to adapt to different structures. And, it is likely using features in only one domain (time or frequency) might provide good results for some specific problems but not for others. In addition, toward an intelligent SHM, its components should be robust to different scenarios, and flexible to various tasks in order to facilitate the incorporation of new events and lower required resources in terms of time, complexity, memory, budget, etc., as much as possible. Therefore, this study develops a practical data-driven method for monitoring the operational state of structures using feature fusion and hybrid deep learning (DL). The proposed method operates on processed data where relevant features extracted from sensory data using signal processing techniques involving AR model, discrete wavelet transform (DWT), and EMD, are integrated into 3-D tensors of features which enter further into a hybrid DL model, called 1-DCNN-LSTM, to identify corresponding structural conditions. The 1-DCNN-LSTM is designed based on the ability to capture local connectivity of the CNN and the well-known performance in accounting for long-term dependencies of long-short term memory (LSTM) network. By doing so, the unobserved underlying patterns at multi-level, multi-scale, and multi-domain embedded in raw structure responses can be extracted and favor damage detection performance. The main contributions of the work are summarized as follows:

- 1) An end-to-end hybrid DL framework is developed for SDD, that includes raw data processing, data augmentation, data fusion, a hybrid 1-DCNN-LSTM model, and damage detection outcomes.
- 2) The correctness and effectiveness of the proposed framework are demonstrated through three case studies, including two experimental and one synthetic database. It achieves comparable performance with 2-DCNN

method while having time complexity reduced by more than 50% and no supplement storage required for images.

- 3) From the proposed method, various studies are conducted to provide insights into the effect of different parameters on SDD performance: 1) the method maintains a good performance when there is data contamination of up to 10% random noise; 2) a reduction in length (10%) of input time-series can lead to a significant decrease in detection accuracy (20%); and 3) increasing the number of sensors effectively improves the damage detection accuracy.

The remainder of this article is organized as follows. Section II briefly introduces various signal processing techniques. Section III details the architecture of the hybrid DL algorithm. Section IV presents the validation of the method via three case studies. Finally, Section V draws conclusions and gives some ideas for future work.

II. FEATURE EXTRACTION FOR SENSOR TIME-SERIES DATA

Acceleration time series obtained from sensor j is denoted by $x_j(t)$, and a system with S accelerometers will provide multivariate time series $X = \langle x_1(t), x_2(t), \dots, x_S(t) \rangle$. It is well known that the distance from sensors to damage location is sensitive to the performance of any SDD method, hence, a SDD method aggregating multiple sensors data will enable to detect potential damages across the whole structure. However, working with multiple time-series data is challenging due to multiple factors such as the curse of dimensionality, long lengths of time series, skewed distribution of measured values, and noise due to device instability, transmission error, and signal distortion. Therefore, it is beneficial to extract underlying features from raw measured data beforehand, then using these extracted features to perform SDD tasks. In the literature, there are a large number of signal preprocessing techniques classified into three groups: time domain, frequency domain, and time–frequency domain, which can capture different unobserved features embedded in time series data. The following subsection describes in detail the techniques adopted in this work.

A. AR Model

AR modeling is a time-domain method used to predict future values of time series data based on previous ones. This technique is adopted here owing to its practicality, effectiveness, and popularity. The number of previous values used in calculation corresponds to the order of the AR model whose formula is expressed as follows [14]:

$$x_{t,j} = \sum_{k=1}^p \phi_{k,j} x_{t-k,j} + e_{t,j} \quad (1)$$

in which the subscript t denotes the time stamp, j stands for the sensor number, p is the order of AR model, $\Phi = [\phi_1, \dots, \phi_p]$ stands for AR parameters, $e_{t,j}$ is the residual error, i.e., the deviation between the estimated value and the

measured one at timestamp j . If $p = 1$, the current value is only affected by the immediately previous value, if $p = 2$, two previous values are needed, and so on. One of the most common ways to determine the AR parameter is the ordinary least squares. Given N signal readings $\langle x_{1,j}, x_{2,j}, \dots, x_{N,j} \rangle$, the AR parameters are estimated to minimize the following square:

$$\Phi_j^{es} = \Phi \left(\sum_{t=p+1}^N [x_{t,j} - \phi_{1,j} x_{t-1,j} - \dots - \phi_{p,j} x_{t-p,j}]^2 \right). \quad (2)$$

In this work, the AR model is determined based on the signal data measured from the intact structural state. Afterward, the obtained AR models are applied to other structural states and the residual errors between the measured values and predicted ones are derived. It is assumed that the nonlinearity caused by damage will lead to significant residual errors, while slight modification of structural parameters only introduces linear variation to the baseline condition, thus, associated residual errors are at a low level. That is why performing SDD on residual errors from the AR model can boost the detection accuracy.

B. Short Time Fourier Transformation

On the other hand, the time-series signal can also be represented in the frequency domain via Fourier transformation (FT). For civil structures, the frequency range of interest depends on both dynamic characteristics of structures and external excitation, which can exhibit time-dependent (nonlinear) behavior. Thereby, the short-time Fourier transform (STFT) technique is developed by scientists, which renders a 1-D time-series in a 2-D image of time–frequency–amplitude representation. Specifically, a long signal is divided into shorter segments of equal length, then FT is applied to each segment. Mathematically, the STFT is formulated as follows:

$$\text{STFT}(\tau, f) = \int_{-\infty}^{\infty} x_j(t) \psi(t - \tau) e^{-2\pi ift} dt. \quad (3)$$

With the time discretization, the formulation is rewritten as follows:

$$\text{STFT}_{n,f} = \sum_{k=n}^{n+N-1} x_{k,j} \psi_{k-n} e^{-2\pi ifk} \quad (4)$$

where $x_j(t)$ is a signal from sensor j to be transformed, ψ is the window function, and N denotes the length of the segment. The 2-D representation of STFT amplitude is called the spectrogram of the signal [15].

C. Discrete Wavelet Transformation

The above STFT technique employed a fixed-width window to extract all frequency contents in time-series, but such a constant window could lead to missing information at lower or higher frequencies. Dealing with low-range frequency requires a long window, whereas, for high-range frequency, a shorter window is needed. To circumvent this drawback, the WT technique is proposed, using varying-length windows

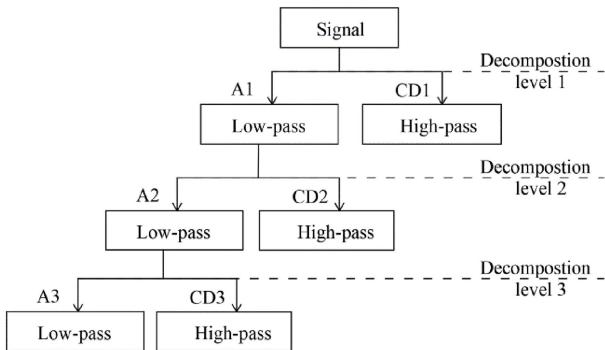


Fig. 1. Schematic representation of discrete wavelet transformation.

to capture better frequency information at different resolutions. Considering a time series $x_j(t)$, the continuous WT CWT_j is defined by the following convolution operator [16]:

$$\text{CWT}(a, b) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{a}} \bar{\psi}\left(\frac{t-b}{a}\right) x_j(t) dt \quad (5)$$

where $\bar{\psi}$ is called mother wavelet, having waveform with finite duration and zero-average, a is a scale variable, and b is a translation variable. Owing to the scaling effect of a , the wavelet function $\bar{\psi}$ has a smaller time duration at high frequency, and longer duration at low frequency, conversely, resulting in the multi-scale resolution characteristic of WT.

For discrete time-series, DWT is used in which the variables a and b are expressed as follows:

$$a = 2^i, \quad b = k \times 2^i \text{ with } i, k \in \mathbb{Z} \quad (6)$$

and the DWT coefficient is written as [5]

$$\text{DWT}_i = \sum_k a_{i,k} \bar{\psi}_{i,k}(t) \quad (7)$$

where $\bar{\psi}_{i,k}(t)$ is discretized wavelets defined by

$$\bar{\psi}_{i,k}(t) = 2^{-i/2} \bar{\psi}(2^{-i}t - k) \quad (8)$$

and

$$a_{i,k} = \int_{-\infty}^{\infty} x_j(t) \bar{\psi}_{i,k}(t). \quad (9)$$

In terms of algorithm, the DWT can be implemented through the multi-rate filterbank algorithm as schematically shown in Fig. 1. There are multiple decomposition levels, each of which contains a half-band low-pass and high-pass filter, namely $h(n)$ and $g(n)$. The signal is fed into the high-pass filter for the analysis of the high-frequency component, and low-pass filter for low-frequency ones. The above decomposition procedure can be repeated multiple times until a small number of samples is left. The original time series can be recovered by using the inverse DWT process and specified high-pass and low-pass filters.

D. Empirical Mode Decomposition

Although the WT can address the locality of signal processing, prevailing patterns of signals could evolve over time, hence the use of a fixed wavelet mother could miss certain

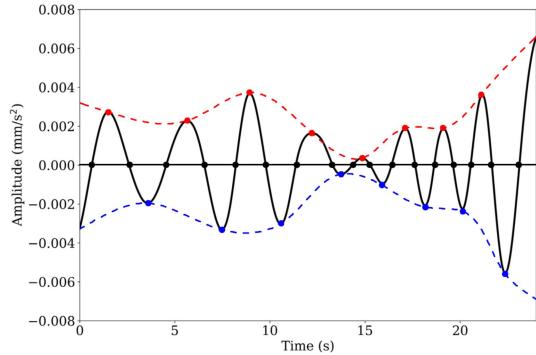


Fig. 2. Example of an IMF with its zero-mean in solid line, maxima, and minima envelopes in dashed lines.

underlying features. To address the nonlinearity and nonstationarity of signals, Huang *et al.* [17], introduced an adaptive method called EMD, which can provide insight into time evolution of data property. To do this, the EMD decomposes the original signals into a finite number of functions called IMFs. An IMF must meet two required conditions: 1) the number of extrema, i.e., minima and maxima, must be equal or differ at most by one that of zero-crossings and 2) the mean value of the upper envelope of IMF formed by connecting local maxima and the lower envelope defined by the local minima is zero at any point as shown in Fig. 2.

Technically, the EMD method is realized by repeating the sifting process involving five following steps. First, the minima and maxima of the signal are identified. Second, two envelopes of the signal are created by using a cubic or higher-order interpolation connecting these minima and maxima. Third, the mean spline of these two envelopes is deduced. Next, the extrapolation is applied to relieve the swing effect at the ends of the signal where extrema values may not be available. Finally, the original signal is subtracted by the mean spline obtained from the third step. If the remaining wave satisfies two aforementioned conditions, then an IMF is obtained, otherwise steps 1–5 are repeated. The sifting process can be performed multiple times to obtain different IMF components until one of stopping criteria is met, for example, a predefined number of IMFs or the remaining signal is smaller than a given tolerance or becomes monotonic, i.e., no extrema can be identified. At the end of the decomposition procedure, one obtains a set of IMFs and a residue. The original signal can be reconstructed by adding all the IMFs and the residue

$$x_j(t) = \sum_{m=1}^M \text{IMF}_m(t) + \text{Res}(t) \quad (10)$$

where $\text{IMF}_m(t)$ is the m th IMF and $\text{Res}(t)$ denotes the residue.

Raw acceleration data from a sensor are fed into the signal processing component with three algorithms presented above to extract relevant features. For AR, the feature used is the residual error computed as the deviation between original time-series with AR prediction. For DWT, extracted features are DWT coefficients in frequency domain as defined in (7). And for EMD, the used features are extracted IMFs expressed in time-histories as defined in (10). In order to bypass the scale

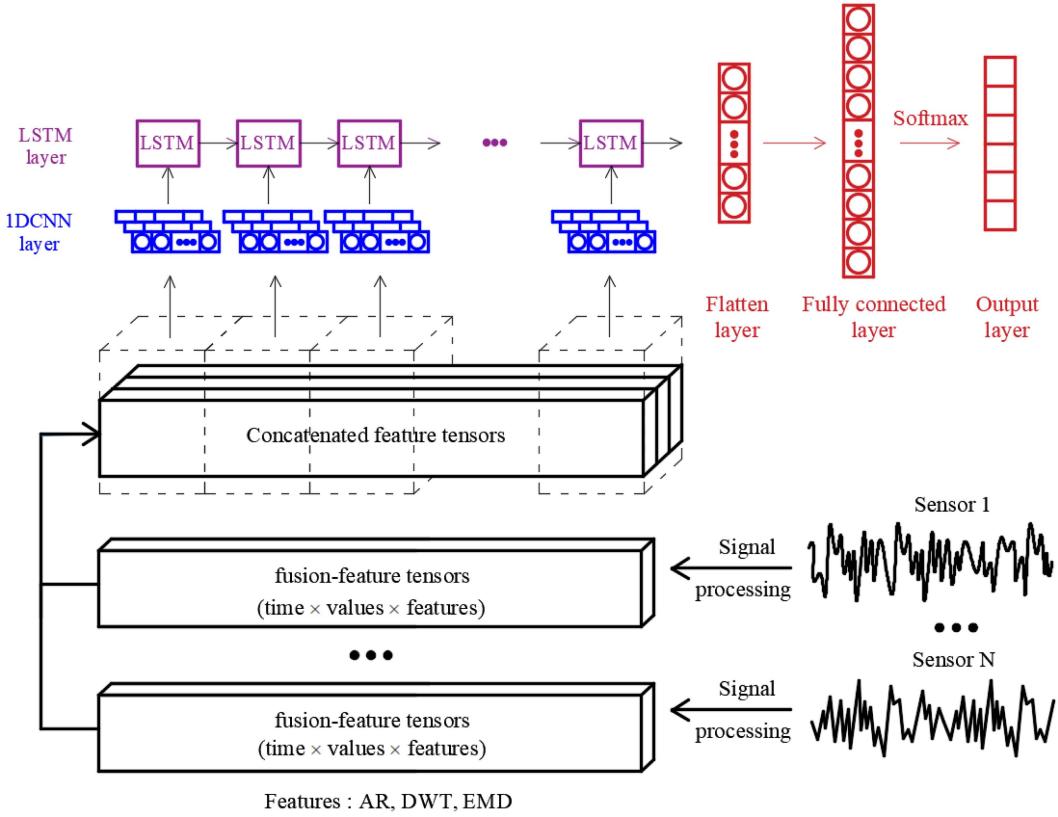


Fig. 3. Architecture of the hybrid 1-DCNN-LSTM network.

difference between features, the standard scaler is applied, resulting in scaled features with zero mean unit variance, which serves as input for the hybrid 1-DCNN-LSTM model. By doing so, one can profit the measurements from all sensors mounted across the structure, favoring the SDD tasks which hardly accomplished by only using a single sensor such as damage localization or damage severity.

III. DL MODEL FOR SHM

The proposed approach is schematically illustrated in Fig. 3, consisting of two main blocks: the preprocessing block whose details are described above, and the hybrid deep neural network which takes as input features processed from the previous block to perform SHM tasks.

A. Hybrid DL Model 1-DCNN-LSTM

It is commonly acknowledged that the CNNs can provide outstanding performance on signal classification and pattern recognition because of two reasons. First, its architecture is especially suitable for discovering local relationships among data points, second, it reduces the number of network parameters, thus leading to a lower computational complexity compared to conventional plain neural network architectures. The formula of one typical convolutional layer is expressed as follows [18]:

$$X^{\text{conv}} = \text{conv1-D}(W^{\text{conv}}, X) \quad (11)$$

where X^{conv} , W^{conv} , are, respectively, the output vector, weight matrix of the convolution layer, X is sensors input, and conv1-D is the 1-D convolution operator. The essential hyperparameters of the convolution layer is the number of kernel N_k signifying the number of local features extracted and the length of kernel L_k denoting the number of surrounding data points are aggregated.

Next, X^{conv} is fed into a LSTM layer, which uses information at multiple previous time steps to perceive insight into recent time steps, referred to as “long-term dependencies” as shown in Fig. 3. The fundamental theory of the LSTM can be found in the work of Hochreiter and Schmidhuber [19].

Introducing \mathcal{L} a typical linear transformation of a combination of X_t^{conv} with N_k features at time step t and an output of hidden layer h_{t-1} with N_h features at previous step as follows:

$$\mathcal{L}(h_{t-1}, X_t^{\text{conv}}) = W[h_{t-1}, X_t^{\text{conv}}] + b \quad (12)$$

where W and b denote weight matrix and bias vector, it is noted that the number of features of \mathcal{L} is equal to that of hidden output h . Each cell of LSTM consists of three gates, namely forget gate f_f , input gate f_i , and output gate f_o , which involve applying the nonlinear sigmoid function σ to a linear transformation \mathcal{L} as below:

$$\begin{aligned} f_f &= \sigma(\mathcal{L}_f(h_{t-1}, X_t^{\text{conv}})) \\ f_i &= \sigma(\mathcal{L}_i(h_{t-1}, X_t^{\text{conv}})) \\ f_o &= \sigma(\mathcal{L}_o(h_{t-1}, X_t^{\text{conv}})). \end{aligned} \quad (13)$$

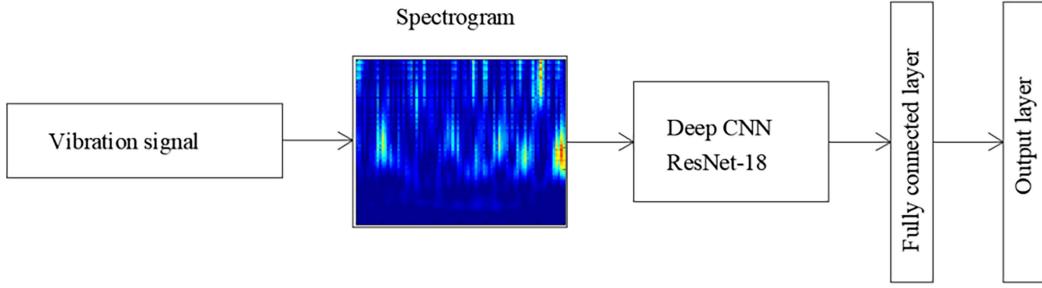


Fig. 4. Workflow of the 2-DCNN-based method for SDD.

On the other hand, a new candidate of information created at time step t is calculated by applying the tanh activation function to a linear transformation of the concatenation $[h_{t-1}; X_t^{\text{conv}}]$:

$$C_t = \tanh(\mathcal{L}_c(h_{t-1}, X_t^{\text{conv}})). \quad (14)$$

Then, the candidate enters LSTM cells:

$$s_t = (f_f \odot h_{t-1}) \oplus (f_i \odot C_t) \quad (15)$$

and a hidden output of the LSTM cell at time step t is calculated at the output gate as follows:

$$h_t = f_o \odot s_t. \quad (16)$$

In these equations, \odot and \oplus stand for componentwise multiplication and addition of two vectors, respectively.

Having established the convolutional layer and LSTM's memory cell, the hybrid DL architecture is schematically illustrated in Fig. 3. Once vibration data enter into the network, it is divided into fixed-length segments, then the 1-DCNN layer will extract local relationships between data points and their surrounding points before feeding to the memory cell of LSTM where long-term dependencies are identified and retained over time. The output of the last LSTM cell will be flattened and fed into a fully connected (FC) layer before being passed to the output layer with the softmax activation function to provide damage identification results. In this hybrid DL architecture, the essential hyperparameters which need to be determined further are the number of kernels N_k , the kernel length L_k in the convolution layer, and size of hidden output N_h at each LSTM cell. The present hybrid DL algorithm is implemented with the help of the open-source machine learning library Pytorch [20].

B. 2-D Convolutional Neural Network

The deep 2-DCNN utilizes as input data 2-D representation of time series. The vibration signal is translated into the time-frequency representation through STFT, as presented in Section II. Then a powerful image classifier in the Artificial Intelligent literature is applied to detect damaged structural cases. Here, one adopted the ResNet-18 classifier [21] owing to its competitive performance, and explicit feed-forward architecture and used the transfer learning technique to finetune the model for structural damage identification tasks. The transfer learning technique allows exploiting general knowledge pretrained with a vast amount of data of

different categories in a particular domain. For further details of the transfer learning technique, refer to [22]. Herein, the parameters of the ResNet-18 model are pretrained with millions of images from the ImageNet data set from Google.

The workflow of the algorithm is illustrated in Fig. 4. First, original vibration signal is fed into a spectrogram pre-process module to be converted into images of time-frequency representation, then entering the ResNet-18 models. Next, data continuously go through two FC layers before giving the classification results at the output layer using softmax activation function. The 2-DCNN is implemented with the help of the DL library Fastai [23].

IV. CASE STUDIES

A. Case Study 1: Laboratory Data

In this section, the hybrid DL structure is applied to a study case involving experimentally measured vibration data from a three-story frame structure realized at Los Alamos National Laboratory [24] as shown in Fig. 5. The data set is selected because of its validity, clarity, and an appropriate number of available data. The frame consists of columns with 17.7 cm length and $2.5 \times 0.6 \text{ cm}^2$ cross section, and plates with 2.5 cm thickness and $30.5 \times 30.5 \text{ cm}^2$ area. These structural components are made from aluminum and joined together using bolts. An electrodynamic shaker at the base floor serves to excite the structure randomly, the excitation is band-limited in the range of 20–150 Hz. At the top floor and the third floor, an additional column ($15.0 \times 2.5 \times 2.5 \text{ cm}$) and a bumper are installed, respectively. The contact between these two elements when the frame vibrates will induce nonlinearity into the dynamic behavior of the frame. Each floor of the structure is equipped with an accelerometer of 1000 mV/g nominal sensitivity to measure the structure vibration. An acceleration signal is recorded for 25.6 s with a sampling frequency of 320 Hz. As the maximum excitation frequency is 150 Hz, then such sampling frequency is large enough to capture essential information content in the structure response.

The above default configuration of the structure is considered as the baseline condition. Afterward, a number of modifications are introduced to the structure to generate different structural state conditions. The modifications involve reducing 12.5% stiffness of one or two columns at each story, adding 19% extra floor's mass at the base or the first floor, and inducing contact between the suspended column at the top floor with the bumper. The structural states not

TABLE I
STRUCTURAL STATE CONDITIONS OF THE THREE-STORY FRAME STRUCTURE

State	1	2	3	4	5	6
Condition	0	0	0	0	0	0
Description	Baseline	Added mass	Added mass	Column stiffness reduction	Column stiffness reduction	Column stiffness reduction
State	7	8	9	10	11	12
Condition	0	0	0	1 (minor)	1 (medium)	1 (medium)
Description	Column stiffness reduction	Column stiffness reduction	Column stiffness reduction	0.2mm gap	0.15mm gap	0.13mm gap
State	13	14	15	16	17	
Condition	1 (medium)	1 (major)	1 (minor)	1 (minor)	1 (minor)	
Description	0.10mm gap	0.05mm gap	0.2mm gap, added mass	0.2mm gap, added mass	0.1mm gap, added mass	



Fig. 5. Three-story frame structure experiment [24].

involving contact are numbered from 1 to 9, and classified as undamaged states. Otherwise, structural states involving the intermittent contact between the column and the bumper are numbered from 10 to 17 and treated as damaged conditions. It is noteworthy that by varying frequency of contact between these two elements through their initial distance, one could generate different levels of damage in the structure (minor, medium, or major). Table I lists all 17 structural states with detailed descriptions. Following the working flow in Fig. 3, measured accelerations from four floors of the frame are processed, then fused into a global 3-D tensor before entering the hybrid network. Note that, here and throughout, a structural state is considered as an output class of the classification problem.

Applying the foregoing processing data techniques to the laboratory data, one can distill raw arbitrary acceleration signal into a set of canonical features which can benefit the damage detection of the structure. Figs. 6–8 show representative results extracted from an acceleration signal at the top floor using the AR model, DWT, EMD, and STFT, respectively. To be specific, the parameters for each technique as follows: the order of AR model is set to 30 [24], the “haar” wavelet family is used for DWT [5], the sifting process for EMD can be found in [24] and for STFT, the length of windowed segment is set to 1000 which ensures to capture available details of vibration signal sampled at the frequency $f_s = 322.58$ Hz.

After applying the AR technique, one obtained a vector of residual error in the time domain with a length of 8192, which was reshaped to a 3-D tensor with a shape of (1, 1, 8192) for the feature fusion later, where 8192 was the original length of the acceleration signal. With DWT, one obtained a vector of DWT coefficients in the frequency domain, as shown in the bottom of Fig. 6, which was interpolated and reshaped into the same shape of (1, 1, 8192) for facilitating the fusion with time domain features in the numerical sense. Using EMD resulted in ten IMFs (Fig. 7), which were combined in a 3-D tensor with a shape of (1, 10, 8192). By fusing features extracted from AR, DWT, and EMD one received a 3-D tensor of (1, 12, 8192) for one sensor, after that features from four sensors are concatenated, resulted in a global 3-D tensors of (4, 12, 8192) before feeding to the hybrid network.

1) *Data Augmentation*: In this subsection, the process of generating data for the hybrid DL is presented. In general, a large and well-balanced database favors the performance of DL algorithms, therefore data augmentation techniques are adopted to increase the size of the experimental data. In principle, the data augmentation technique introduces some minor changes in the original data without altering its underlying pattern. Herein the utilized techniques are flipping (rotation), scaling, permuting [25]. Flipping inverts the sign of the signal, scaling increases/decreases the magnitude of the raw data slightly by a random ratio from 5% to 10%, and permuting will swap two randomly selected small fractions (2% length) of the signal. Fig. 9 illustrates how data augmentation techniques work. After applying data augmentation techniques, the size of the final database increases up to 1000 time series, which is sufficient for training and validation of the proposed hybrid DL model.

2) *Data Preparation*: Formally, the database is split into three subsets, i.e., training, validation, and test one with a user-defined ratio. However, a single split might not ensure a well-balanced distribution of different structural conditions among subdata set. Therefore, the k -fold cross-validation strategy is employed to reduce the bias of the model. First,

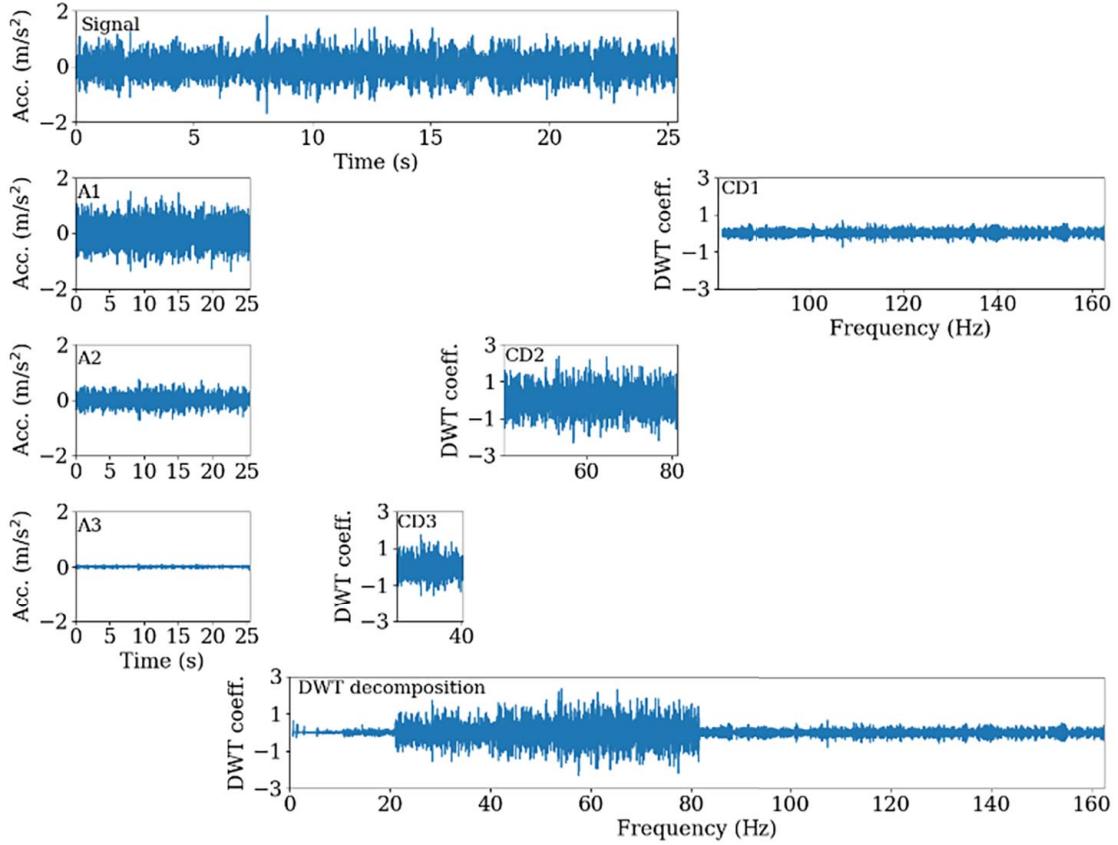


Fig. 6. DWT of a vibration signal. The first row is the original signal, the second to fourth rows are three levels of DWT decomposition where leftmost figures are remaining components after each DWT decomposition level, and rightmost figures are resulting DWT coefficients. The last row is the total DWT transform of the original signal.

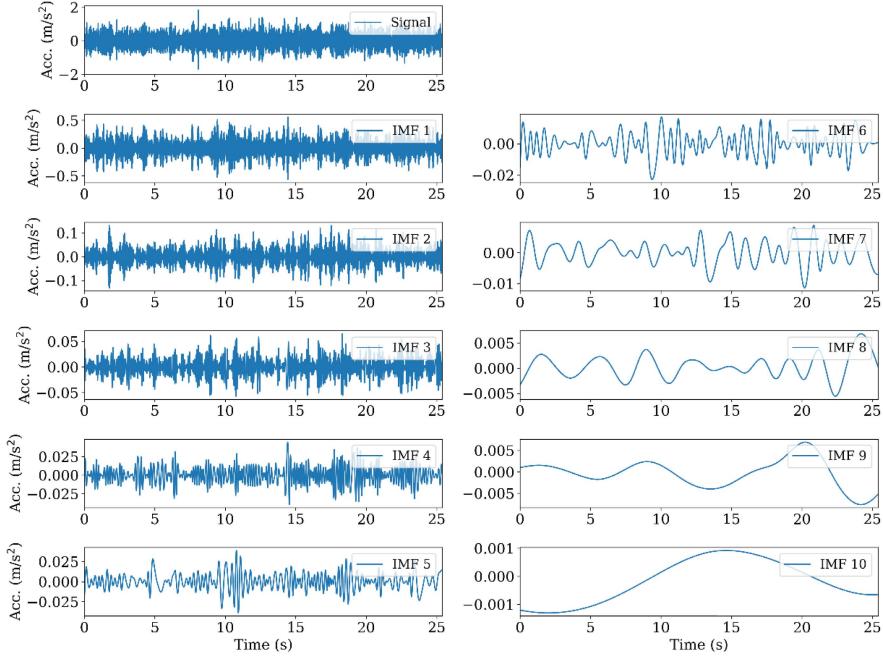


Fig. 7. EMD of a vibration signal. The first row is the original signal, the next rows are resulting IMFs. There are in total ten IMFs for this example.

the whole database is divided into training (90%) and test data (10%). Second, the training data are augmented with the presented data augmentation technique, before being split

into the K equal portions. Here, a common value $K = 10$ is selected [26], meaning the training process will be iterated ten times, each time one different portion is used for validation,

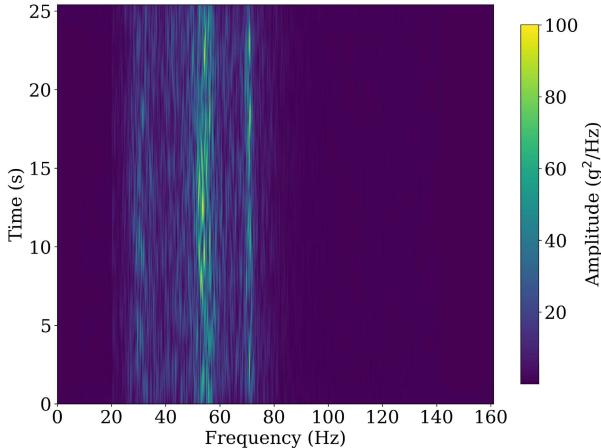


Fig. 8. Spectrogram of a vibration signal obtained by using the STFT. The brighter the color, the higher the amplitude, meaning that more signal energy concentrates over corresponding frequency ranges.

whereas the remaining serves for training as illustrated in Fig. 10. Once the model is trained, its final performance is reported on the test data unseen during training. To be specific, the size of test data and augmented training data are 60 and 540 samples, which results in global feature tensors with the shape of (60, 4, 12, 8192) and (540, 4, 12, 8192), respectively.

3) *Training Process*: As previously mentioned in Section III, the hyperparameters of the proposed hybrid architecture are the number of kernels N_k , the kernel length L_k , in the convolution layer, and the number of hidden layers N_h in LSTM cell. Actually, there does not exist a common way for the selection of the best parameters, but it depends on specific problems. Thus, one adopts the grid search technique to test all possible combinations of hyperparameters for the identification of the optimal architecture. Specifically, k varies in the range [5, 50], L_k in [10, 100], and N_h in [3, 30]. It is noted that other parameters are set by common values found in the literature and fixed throughout the whole training process. Specifically, the optimizer is Adam [27], the learning rate standing for the relative amount of DL model weights updated after each optimization iteration, is set to $lr = 0.0001$, the number of epochs $N_{\text{epoch}} = 200$, the batch size, i.e., the number of data utilized in one optimization iteration is 32, and stop training early with a patience value of 20, meaning the training process will be stopped if no improvement is observed after 20 consecutive iterations. These values are defined to ensure covering as many details as possible in behaviors of DL model during the training process. It is recorded that the final hybrid DL model with the number of convolutional kernels $N_{\text{conv}} = 50$, the kernel length $L_k = 200$, and the number of LSTM cell's hidden layers $N_h = 20$ provides the highest averaged accuracy of 93.5% and a standard deviation of 1.4% on the validation data set.

4) *Damage Detection Results*: Table II compares detection results obtained by each method, in which the first row presents the mean value of validation accuracy of tenfold technique, the second row is the standard deviation, the last two rows are the used resources, i.e., CPU time and memory. Note that the calculated result with raw data is presented here

TABLE II
COMPARISON RESULTS OF DAMAGE DETECTION PERFORMANCE BETWEEN MODELS

Model	Raw	DWT	AR	EMD	Fusion	2DCNN
K-fold mean (%)	81.4	90.2	93.2	88.1	93.5	93.7
K-fold std (%)	6.6	2.5	1.2	1.5	1.4	1.2
CPU time (min)	26	28	26.5	41	45	118
Storage (Mb)	49	49	49	49	49	292

TABLE III
COMPARISON RESULTS OF DAMAGE DETECTION PERFORMANCE BETWEEN MODELS PERFORMED ON NOISY DATABASE

Model	Raw	DWT	AR	EMD	Fusion	2DCNN
K-fold mean (%)	80.1	86.1	68	86	92.1	91.2
K-fold std (%)	5.9	2.1	3.8	1.7	1.8	1.5

for the comparison purpose, obtained by entering the raw data from sensors directly to the concatenate layer then the hybrid 1-DCNN-LSTM, bypassing the preprocessing blocks (Fig. 3). As observed, the 2-DCNN provides a high accuracy result of 93.7% on the validation data set, which is approximately achieved by the proposed feature-fusion method. The hybrid DL model using raw time-series data yield significantly lower accurate results, i.e., 81.4%, whereas models with DWT and EMD provide better results, i.e., 88.1% and 90.2%, respectively. However, the method using only the popular AR technique can also nearly attain the same accuracy as the sophisticated 2-DCNN and the feature-fusion technique. This confirms the usefulness of the popular AR technique, although simple but efficient in some scenarios. On the other hand, the required CPU time for the training process associated with each method clearly shows the high demanding resources from 2-DCNN, both in time and storage.

In practice, measured vibrational data encompass inevitable noise related to data acquisition units, device instability, and so on. Therefore, it is of great importance to quantify the robustness of a SHM application in dealing with noisy data before applying to the real-world structure. It is noted that the data considered above are acquired in controlled laboratory conditions though some small amount of noise is also present, in the next step, the performance of methods when dealing with a more pronounced noise is explored. The white-noise is defined by the following equation:

$$X_{\text{noise}}(t) = X(t) + \alpha \times \eta(t) \quad (17)$$

in which $X(t)$ and $X_{\text{noise}}(t)$ are original and added-noise time series, respectively, $\eta(t)$ is a white noise time series with zero mean and unit variance, α is the noise amplitude compared to the rms value of the signal amplitude $X_{\text{rms}} = ((X_1^2 + \dots + X_n^2) \times 1/n)^{1/2}$, where X_i is a value of the signal at time instant t_i . The white noise reflects the avoidable unknown ambient excitation, whose amplitude α can be set to 10% on average as considered in [28] and [29]. Table III and Fig. 11 present in detail performance results for methods of interest on noisy data. It is noticed that the performance of the hybrid DL using only AR coefficients decreases drastically with added noise, from about 90% down to below 70%, while the 2-DCNN

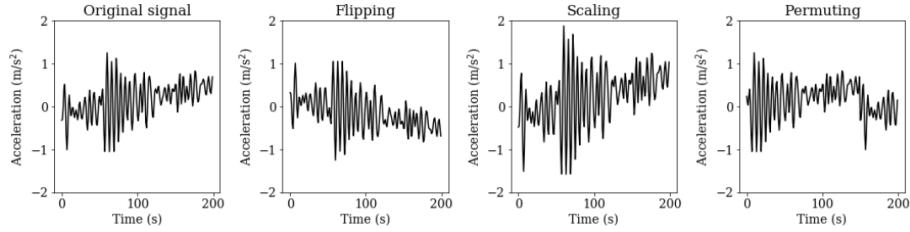


Fig. 9. Data augmentation techniques for time-series data.

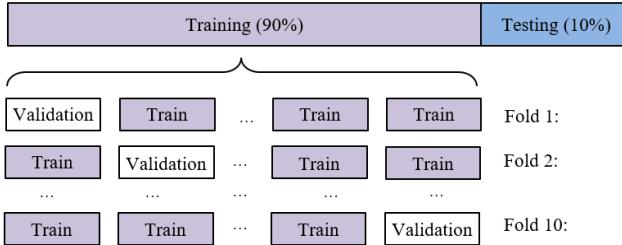


Fig. 10. Tenfold cross-validation strategy.

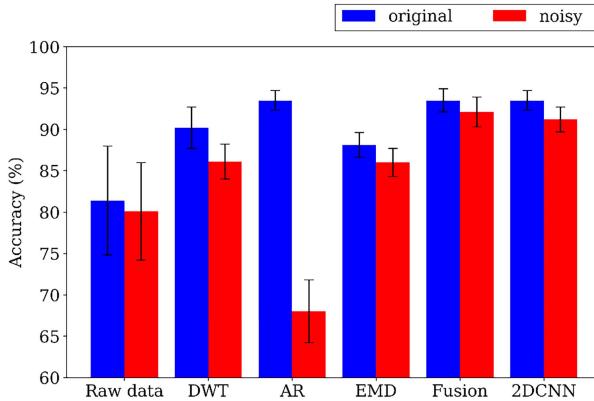


Fig. 11. Comparison of SDD obtained by different models. The fusion method and 2-DCNN provide the best results.

and fusion-feature method can retain sufficiently high accurate results of 91%. The observed decrease in performance of methods using DWT and EMD is also significantly lower than that of AR, i.e., a reduction of 4.0% and 2.1%, respectively. This can be explained by the fact that the incorporated white noise has flat broadband frequency spectra in nature, hence the feature extraction method in the frequency domain and time-frequency domain such as DWT, EMD, and STFT still discern the underlying frequency components from the noise. Indeed, there are methods using AR coefficients, which are able to address noisy data effectively, for example, the works in [30] and [19] extracting modal characteristics of the structure from AR coefficients. However, these methods require further mathematical transformation and expert knowledge in the dynamics of structures. Besides, to ease the fusion with other signal processing techniques, residues obtained from AR models in a time-series form is preferable than tabular data of frequency values or mode shape vectors, in this work.

In short, these above results imply that fusing advantages of different feature extraction techniques favor the hybrid

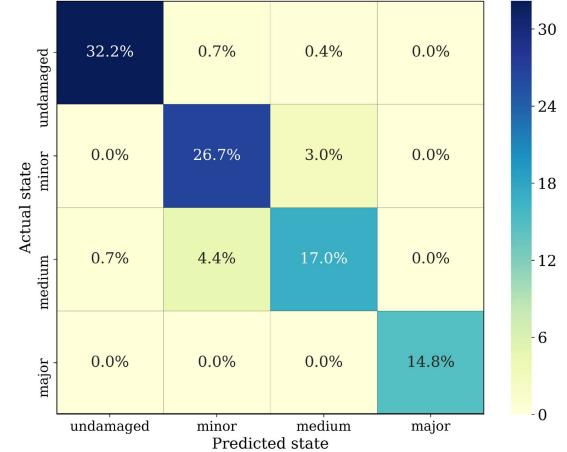


Fig. 12. Confusion matrix for damage severity classification. Resulting global classification accuracy is 90.7% by summing diagonal terms.

DL method's robustness and reliability when encountering different scenarios in reality.

5) Damage Severity Results: As shown in Table I, the damaged state of the structure can be divided further in order as minor, medium, and major damage. Then, the proposed method is adapted to address the damage severity task by fine-tuning the numbers of perceptron at the output layer to four and relabeling the database by corresponding damage levels. Note that it is not a trivial task as the difference between the minor damage state with the pristine one, or between two consecutive damage levels, is subtle and can be overlapped by random noise, that is why the required classifier should possess robust discriminative power.

After carrying the training stage with tenfold cross-validation, the network is tested on the unseen and also noisy data. Note that the network architecture presented previously is kept unchanged, i.e., $N_{\text{conv}} = 50$, $L_k = 200$, $N_h = 20$. Fig. 12 illustrates the detection results via a confusion matrix where the diagonal terms denote the correctly detected damage cases in terms of global percentages, the off-diagonal terms for false detection, and all values of the matrix sum up to 100%. It is ideal that all nondiagonal values are zeros. Thus, by summing the diagonal terms, one obtains the global classification accuracy of the proposed method as high as 90.7%. These results reaffirm the capability in capturing the underlying dynamical features and mapping them to the corresponding structural states of the featured-fusion approach.

Next, one investigates the impact of the effective length of the signal on the detection performance. The same process

TABLE IV
LABELING FOR DAMAGE LOCALIZATION TASK OF THE THREE-STORY FRAME STRUCTURE

State	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Location	S0	S0	S0	S1	S1	S2	S2	S3	S3	S4								

S0: No damage. S1, S2, S3: stiffness reduction of columns at the first, second, and third story, respectively. S4: intermittent contact at the fourth story.

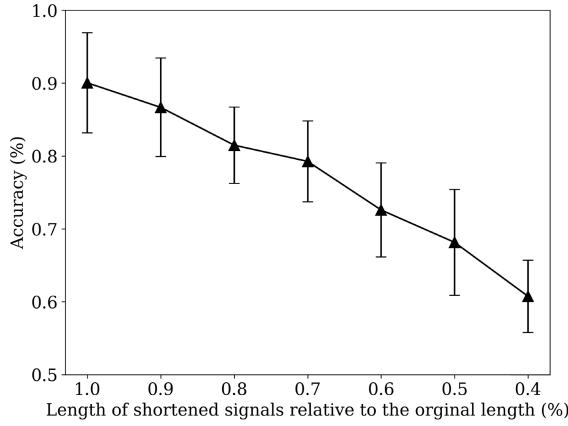


Fig. 13. Damage severity performance versus the length of shortened signals.

with the damage severity task is carried out except for the length of the signal, which varies from 40% to 100% of the original one. For instance, a 50% shortened signal means only the first half of the signal is taken as input for the proposed approach, another half is removed. The evolution of the detection accuracy in a function of the relative length of the shortened signal is depicted in Fig. 13. As expected, there is a tradeoff between the performance and the length as a clear downward trend is observed, for example, a 20% reduction in length leads to 10% decrease in accuracy (from 90% to around 80%).

6) *Damage Localization Task:* As demonstrated, the feature-fusion hybrid DL method can perform the level 1 of SSD, i.e., damage detection with high accuracy, then in this paragraph the present method is adapted to perform the level 2 of SSD, i.e., damage localization. To this end, the experimental database is relabeled, each time-series is annotated by the story where modification is realized accordingly, as listed in Table IV. Structural states 1 through 3 are labeled as S0, meaning no damage is introduced, states 4, 5 as S1, states 6, 7 as S2 and states 8, 9 as S3, since the column stiffness at the first story, the second story, and the third story are reduced, respectively. States 10 through 17 are labeled as S4 because the intermittent contact caused by the additional short column happens at the fourth story. Next, a similar procedure involving data generation, data augmentation, training, and validation procedure, as described above, is conducted. Fig. 14 details the localization results for each class, where diagonal terms denote accurate predictions, and off-diagonal terms refer to misclassifications. Summing the diagonal terms yields a global classification accuracy of 91.0%. This result shows that the proposed method is applicable for level 2 of SDD, and the conversion between

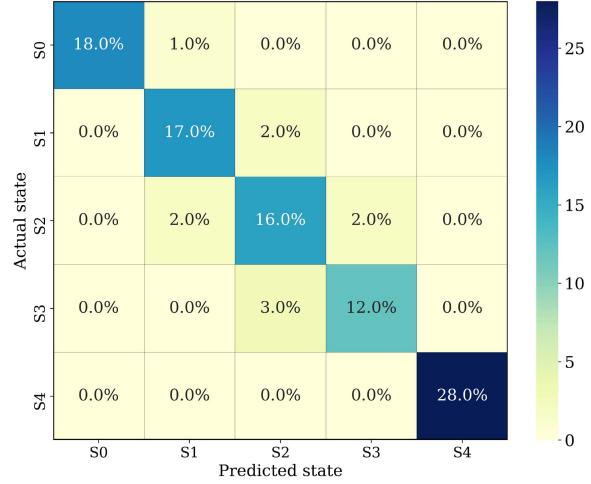


Fig. 14. Confusion matrix for damage localization classification. Resulting global classification accuracy is 91.0% by summing diagonal terms.

different SDD tasks is fairly straightforward without requiring any complex and tedious manual calculations.

B. Case Study 2: My Thuan Stayed-Bridge

For the second case study, one applies the proposed fusion feature DL approaches to a synthetic database of the full-scale bridge My Thuan in Vietnam, which is a stayed-cable bridge playing a vital role in the transport system. Among the most critical structural elements of the bridge, cables have a great impact on the dynamic behavior of the structure. A significant loss in the cable's prestressed force will result in reduced structure rigidity, then causing potential excessive vibration and inducing further damages such as crack, fatigue, and so on. Hence, using the present approach provides an alternative approach to quickly and remotely assess the location of tendons with tension loss based on the dynamic response of the bridge before requiring workers to perform *in situ* tests to check tendons thoroughly.

1) *Finite Element Model:* Fig. 15 presents main geometric parameters of the bridge: the total length of the bridge is 650 m with a 350 central span and two side spans of 150 m, the bridge reinforced concrete deck has a width of 23.6 m, a depth of 0.2 m, and is supported by two longitudinal girders of 2.0 m depth and transverse beams placed at 5.2 m intervals. The two H-shaped bridge towers have a total height of 129.5 m, including legs of 2.5 m width. The materials of the superstructure components are concrete of Grade 50 with $E_c = 35.75$ GPa, a volumic density of 2400 kg/m^3 , compressive strength of $f'_c = 50$ MPa. The system of cables consists of 128 cables with a diameter of 15.2 mm, an elastic modulus of 195 GPa, and a tensile strength $f_{pu} = 1860$ MPa, dividing

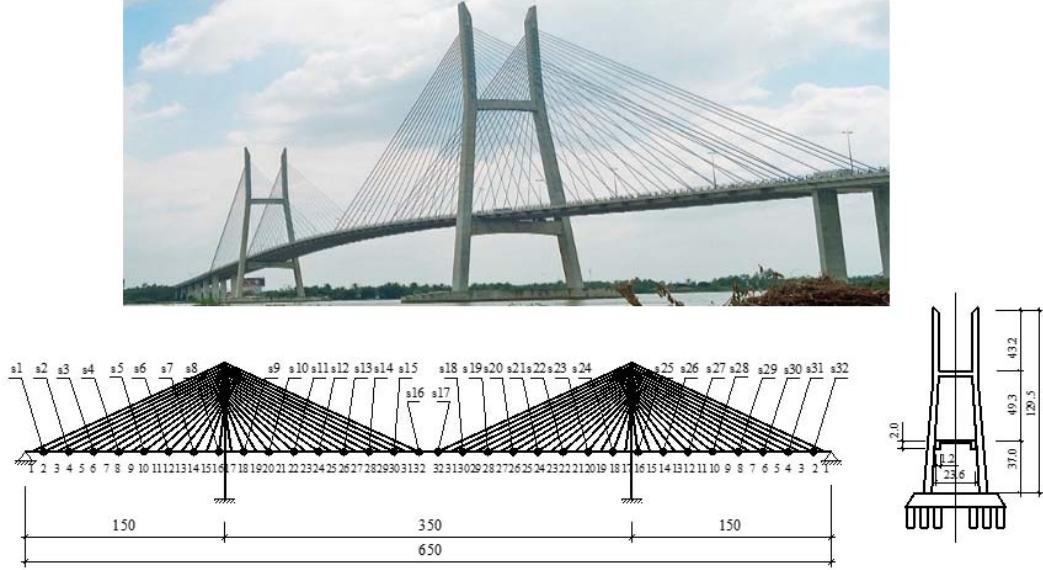


Fig. 15. My Thuan stayed-bridge in Vietnam.

TABLE V
TENSION FORCE IN CABLES

Cable	1	2	3	4	5	6	7	8	9	10	11
Measured Force (kN)	6395	5264	4554	3668	3459	3459	3481	3478	3266	3128	2935
Length (m)	177.5	173.8	170.4	162.4	152.5	142.8	133.1	123.6	114.3	105.2	96.4
Cable	12	13	14	15	16	17	18	19	20	21	22
Measured Force (kN)	2543	2267	2497	2042	2171	2512	2376	1938	2135	2508	2743
Length (m)	88.0	80.0	72.6	65.5	58.5	57.4	63.4	69.6	76.4	83.9	91.9
Cable	23	24	25	26	27	28	29	30	31	32	
Measured Force (kN)	2403	3033	2947	3145	3417	3688	4077	4006	4489	5967	
Length (m)	100.4	109.2	118.3	127.6	137.1	146.8	156.5	165.4	176.3	186.4	

into four groups, each connected to one pylon. As the layout of the cables is symmetric, the groups of cables have similar influences on the bridge's behavior. The cables are annotated from 1 through 32 within each group, which is also the number of output classes of the DL model for this case study. The prestressed force in each cable is provided in detail in Table V, according to as-built design drawings.

Herein, the SDD task is referred to as the damage localization, meaning that the proposed method is used to inversely identify the location of the cable with tension loss based on acceleration time series computed at the points uniformly distributed across the bridge. It is widely acknowledged that acceleration is rich in discriminative features, thus suitable to take as inputs to SDD models [31]. To be specific, there are 32 measured points on each side of the bridge, as illustrated in Fig. 15, resulting in a total of 64 points. The same workflow presented for the first example is carried out, including using AR model, DWT, and EMD to extract salient features in both time and frequency domain from raw acceleration time series and the k -fold validation strategy. However, here one needs to integrate information from a significantly larger number of sensors, which can be efficiently realized by using the convolution layer in the proposed hybrid DL model.

To generate the database for the DL model, one adopts the Monte Carlo simulation upon a calibrated finite element model (FEM) of the bridge using model updating technique and *in situ* measured data. At first, an initial 3-D FEM of the bridge is constructed from the as-built design drawings using the software Abaqus [32], then a series of measurements is carried out to measure the vibration of each cable in service, i.e., under vehicular loads. After that, an optimization-based model updating technique is used to minimize the difference between modal characteristics of cables with the numerical counterpart, including eigenfrequencies and mode shapes. As detailed solutions for similar problems can be found in the optimization literature [33], [34], thus for brevity, one only presents measured forces in cables in Table V.

In terms of the FEM, the Abaqus software is adopted because of its modeling power and practical Python interface programming to build the bridge model. The bridge deck is modeled by four noded S4R shell elements, the pylons and reinforced concrete girders are modeled using 3-D beam elements, the stay cables are modeled using two-noded T3D2 truss elements subjected to prestressed action. The prestressed value per strand is obtained by dividing the tension force by the strand area and is modeled as an initial condition, directly applied to cable elements at the beginning of the

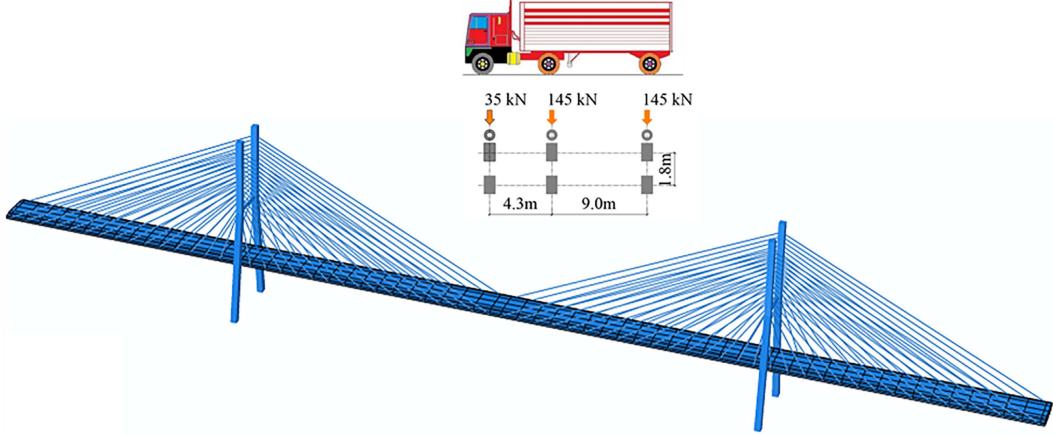


Fig. 16. Finite element model of the stayed-bridge.

TABLE VI
COMPARISON RESULTS OF DAMAGE LOCALIZATION PERFORMANCE
BETWEEN MODELS

Model	Raw	DWT	AR	EMD	Fusion	2DCNN
Accuracy (%)	22.5	28.0	43.0	78.0	84.5	85.2
CPU time (min)	110	135	125	170	180	486
Storage (Gb)	1.3	1.3	1.3	1.3	1.3	2.9

analysis using predefined mechanical fields. In this work, the source of excitation is the vehicular load which is the most common type of load encountered in the bridge structure. The vehicular load is modeled using a standard HS20 truck after the American Association of State Highway and Transportation Officials specifications [35] moving with a constant velocity whose details are highlighted in Fig. 16. In order to enrich the database for training the DL model, the number of cars varies from 1 to 5, the velocity of the cars is in the range of [10, 15, 20, 25, 30] m/s, and various damage scenarios are introduced, involving 100% tension loss at a random cable, resulting in an extensive series of 10000 simulations. Afterward, the vertical accelerations at 64 virtual sensors uniformly spaced along the bridge deck are virtually recorded, and fed into the present approach to identify the respective location of the defected cable. After the data preparation step, the test data and training data have the shape of (1000, 64, 12, 6000) and (9000, 64, 12, 6000), respectively where 64 is the number of sensors, 12 is the number of extracted features, i.e., AR, DWT, IMF, and 6000 is the length of time-series measured for 60 s with a sampling frequency of 100 Hz.

2) *Computation Results:* Table VI compares the obtained detection accuracy using the feature fusion approach with the powerful 2-DCNN, as well as models using separately extracted features. Apparently, the DL-based method can properly perform the damage localization tasks based on acceleration signals. Such a task is not trivial as it requires a powerful discrimination ability because the number of outputs herein is 16 times as much as that of the damage detection task with binary outputs (damaged/undamaged) in the first case study. Furthermore, the proposed approach can equivalently

achieve a considerably high accurate result of about 85% as the 2-DCNN, following by the model using EMD, AR, DWT, and raw data. Fig. 17 shows in more detail the accuracy percentage for each cable through a confusion matrix.

All the calculations were conducted on a high-end machine equipped with 16 cores CPU Intel Xeon, 64-GB RAM, and 2 Geforce GTX 1080Ti GPU. The model using directly raw data requires the least CPU time of 110 min; the DWT and AR have higher CPU time, mainly due to the feature extraction step, as listed in Table VI. For the EMD, as the number of features increases to 10, corresponding to a training data with the shape of (9000, 64, 10, 6000), the CPU time increases to 170 min, and for the model with feature fusion, i.e., the number of features is 12, the CPU is 180 min. As the calculations were realized with the library Pytorch, which allows the parallel computation and take advantage of GPUs; thus, the computational time with fusion-features can be reduced. On the other hand, the factors having considerable influences on the CPU time are the number of data and the length of time-series, that is why the CPU time in this example is significantly larger than those in the case study. In comparison, 2-DCNN demands 490 min for the training process and more than twofold memory volume for respective 2-D representation of each time-series.

In addition, the effect of the number of virtual sensors on the model performance is explored by varying the former in the range of [1, 2, 4, 8, 16, 32, 48, 64]. As can be seen in Fig. 18, a smaller number of sensors leads to significantly reduced detection accuracy. On the other aspect, for this example, the DWT and the raw time-series yield low accurate detection results despite increasing the number of sensors, possibly because the mass of vehicular load is fairly small compared to that of the bridge. Then, the difference between the frequency range of forced vibration responses with free vibration of the bridge is subtle, and limit the capacity of the frequency-domain method. Besides, the time-domain such as the AR model gives better accuracy, and the EMD provides dominant contributions to the performance of the fusion-feature method. Overall, a stable trend is observed when more than 32 sensors are used.

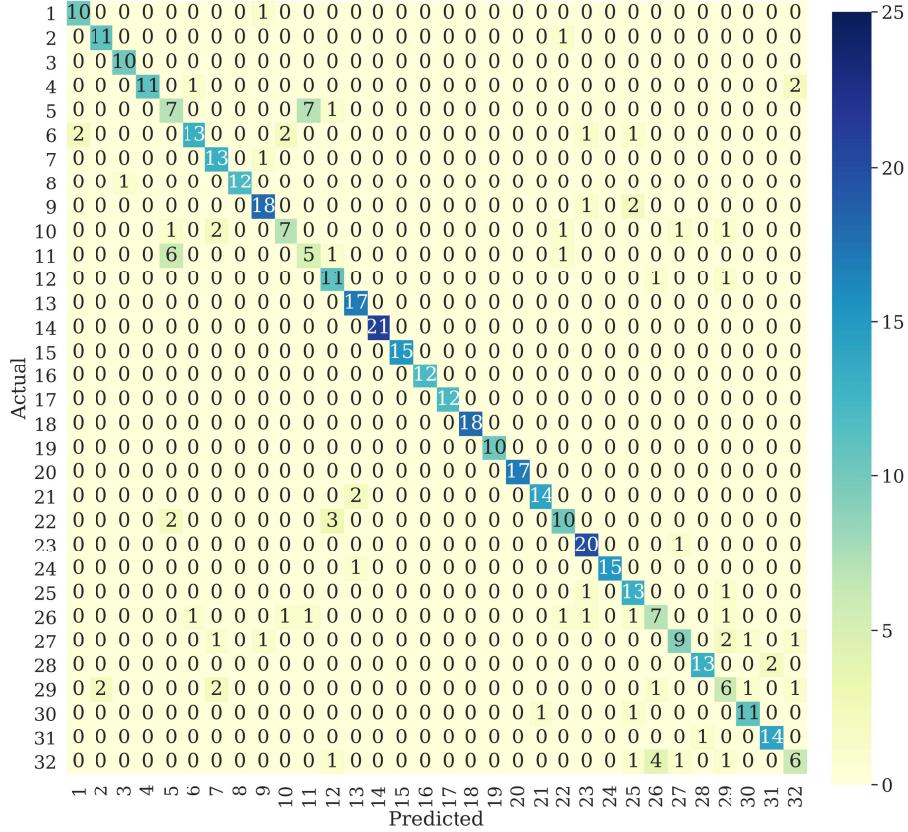


Fig. 17. Confusion matrix results of the damage localization task for the stayed-bridge structure.

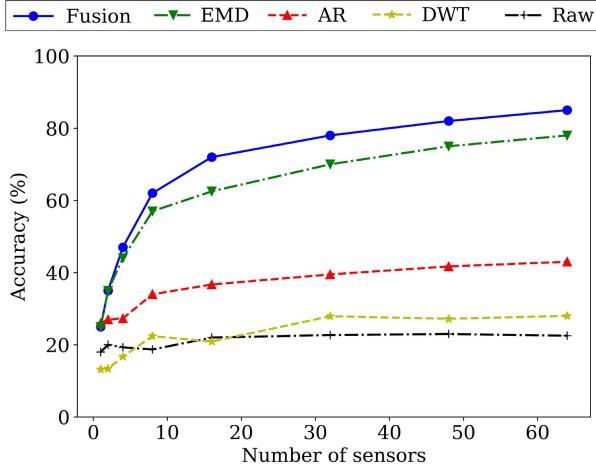


Fig. 18. Model accuracy in functions of the number of sensors.

C. Case Study 3: Progressive Damage Tests of Z24 Bridge

In this subsection, the proposed framework is extended to a real database of the Z24 Bridge in Switzerland from the BRITE-EURAM project SIMCES (System Identification to Monitor Civil Engineering Structures) [36]. The Z24 bridge is a prestressed bridge overpassing a highway connecting Bern and Zurich. The bridge consists of three spans with a total length of about 60 m, as illustrated in Fig. 19. Before the bridge was completely demolished, different progressive damage scenarios were carried out over a month in 2008,

including lowering, lifting of piers, spalling of concrete, failure of anchor heads, and rupture of prestressed tendons. Details of progressive damage tests (PDTs) and corresponding labels are reported in Table VII. A system of sensors was strategically installed across the bridge to measure its structural responses during PDTs. For each PDT, nine test setups were conducted, and for each setup, 33 acceleration time-series from various sensors were collected, of which five time-series, namely R1-V, R2-T, R2-V, R2-L, and R3-V were common between setup tests, where R1, R2, R3 are the location of sensors, V, T, L denote the directions of displacement, i.e., vertical, transversal, and longitudinal.

In order to prepare databases for the DL model, a similar procedure successfully reported in [4] is adopted. More specifically, data from a sensor is sampled with a frequency of 100 Hz, resulting in a signal of 65 568 samples, which is further divided into eight time-series of length 8000. Hence, for each PDT with nine test setups, 72 time-series are extracted from one sensor. Then, the obtained database of Z24 consists of 1224 groups of five time-series, each group labeled from 1 to 16 according to PDT associated. Applying the data preparation step results in a test data and augmented training data with the shape of (231, 5, 12, 8000) and (4769, 5, 12, 8000), respectively where 5 is the number of sensors, 12 is the number of extracted features, i.e., AR, DWT, IMF, and 8000 is the length of time-series.

Next, the tenfold cross-validation strategy is applied to train and validate the detection accuracy of the proposed approach. The key parameters of the DL architecture after fine-tuning for this

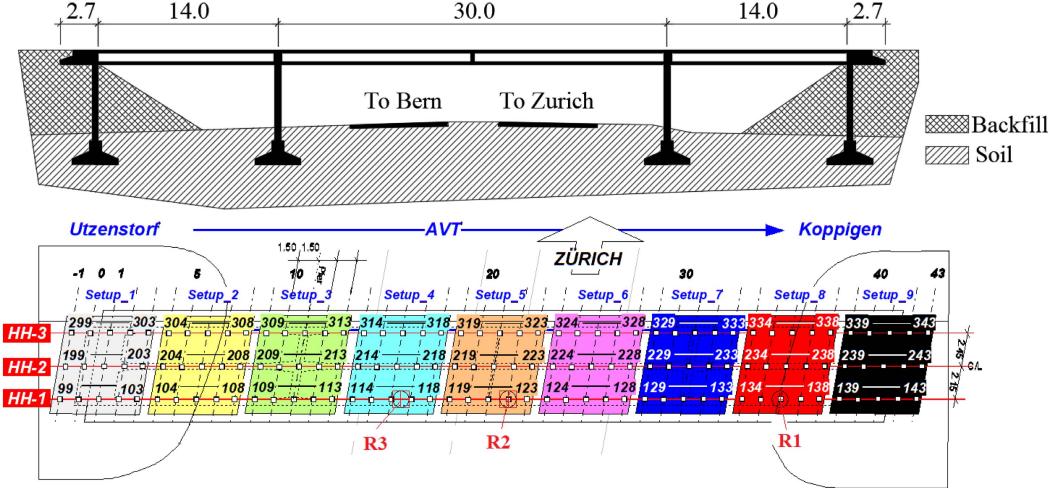


Fig. 19. Z24 bridge in Switzerland: the geometry of the bridge is depicted at the top, locations of sensors are presented at the bottom.

TABLE VII
LABELING FOR THE SDD TASK OF THE Z24 BRIDGE

State	1	2	3	4	5	6	7	8
Description	No damage	No damage	20 mm settlement	40 mm settlement	80 mm settlement	95 mm settlement	Tilt of foundation	No damage
State	9	10	11	12	13	14	15	16
Description	12 m ² spalling	24 m ² spalling	1 m landslide	1 column failure	2 anchors failure	4 anchors failure	2 tendons failure	4 tendons failure

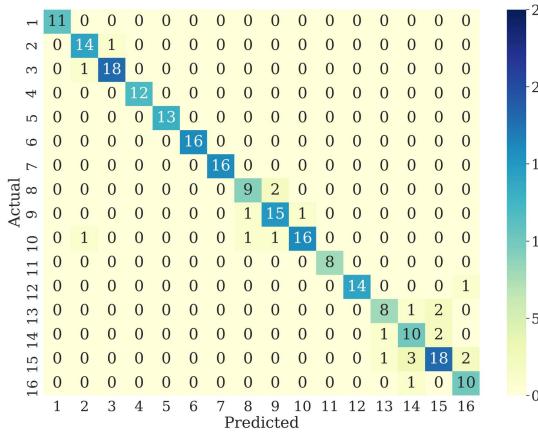


Fig. 20. Confusion matrix for SDD results of Z24 bridge.

study-case is as follows: the number of convolution kernels $N_k = 200$, the length of convolution kernel $L_k = 100$ and the size of hidden output of LSTM cell $N_h = 100$. Table VIII summarizes the damage detection results obtained by the fusion-feature method and their counterparts using one extracted feature. It can be seen that the fusion-feature method achieves a highly accurate SSD result of 90.1%, more detailed classification result for each PDT is illustrated via the confusion matrix in Fig. 20, moreover, PDTs in same categories are also clearly separated, for example, PDT 9 and PDT 10 for concrete spalling, PDT 15 and 16 for tension failure, which are sometimes misclassified by other methods [4].

TABLE VIII
SDD RESULTS FOR PDT OF Z24 BRIDGE

Model	Raw	DWT	AR	EMD	Fusion
Accuracy (%)	61	78.6	70.6	81.2	90.1

To provide more intermediate results about the performance of the hybrid model, the t-distributed stochastic neighbor embedding (t-SNE) [37] method is employed to show the clustering of different conditions of the structure. It is expected that measurements from a given structural condition will be grouped into the same cluster and well separated from others. In contrast, if data points are mixed together, meaning further improvement is required to achieve a better SDD performance. Fig. 21 illustrates the t-SNE representation on test data obtained at different main steps of the hybrid 1-DCNN-LSTM, namely, after the 1-DCNN step, the LSTM step, and the FC step. It can be seen that at the 1-DCNN step, data points from different classes overlap together, no observable clustering is obtained, while after the LSTM step, the clustering is much better, data points are divided into different groups, though some classes are not clearly separated. At the FC step, data points are almost split; only a small number of data points are mixed at their corresponding class boundaries; for example, those from classes 14 and 15.

In short, the validity of the present method is successfully proved through experimental data from a real-world bridge.

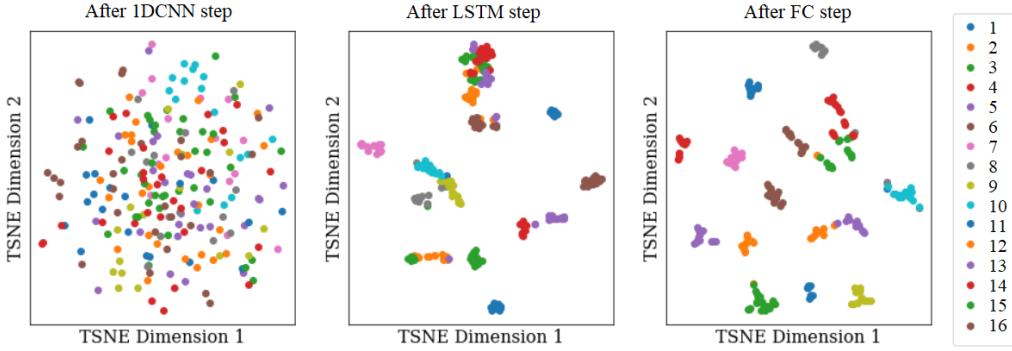


Fig. 21. t-SNE visualization of damage localization task obtained at main steps of the hybrid 1-DCNN-LSTM algorithm.

V. CONCLUSION

In this work, a novelty method for SDD was developed based on the feature fusion technique and a hybrid DL architecture for improving the detection performance with low required resources in terms of computational cost and storage capacity. Remarkably, the present method works on pure data; thus, a numerical model of the structure is not required, which allows addressing real-world problems in which pure forced excitation is challenging to identify. Moreover, the method is able to perform not only level 1 of damage assessment but also level 2 (damage localization). Such a SDD method enables to monitor structures in a smart and real-time fashion.

The applicability and efficiency of the proposed method were first validated through a case study with experimental vibrational data obtained from a three-story frame. Two levels of SDD were conducted, i.e., damage detection and damage localization, with both original data and data contaminated by random noise. It was found that the feature fusion method outperforms the other ones using separately features in terms of accuracy and robustness. When comparing with the sophisticated 2-DCNN method, the proposed one provided equivalent performance. Moreover, it significantly reduced both time and memory complexity, which would be costly when performing long-term monitoring and when structures become more complex. Second, real data from PDTs of the Z24 bridge were used to support the validity of the fusion-feature method. It showed that the proposed method consistently achieves highly accurate damage detection results, while performances of methods using separately extracted features could vary significantly depending on specific SDD tasks. Next, the method was applied to the full-scale cable-stayed bridge My Thuan in Vietnam for the detection of tension loss in prestressed cables. The case study emphasized the discrimination power of the proposed method in the detection of structural damage when dealing with a multi-classification problem of a complex structure using simultaneously multiple time-series input. Moreover, the low time complexity and memory usage of the proposed method facilitated parametric studies, which quantified the influence of the number of sensors on the detection accuracy, providing tradeoff constraints for engineers and owners when drawing an optimized maintenance strategy.

In future works, it is noteworthy to extend the method to an online framework where measured data are collected from Internet of Things (IoT) sensors, then uploading to cloud

servers. After that, the present SDD method can analyze data and provide informative structural assessments for responsible personnel from anywhere in the world through a web application. Another exciting direction is to incorporate probabilistic analysis into DL frameworks, then even with limited input data, one is still able to perform SDD tasks with corresponding confidence intervals. When more data are available, the model will be updated, and the variability of the detection results will reduce, i.e., the confidence interval will become narrower.

REFERENCES

- [1] C. N. Macleod, G. Dobie, S. G. Pierce, R. Summan, and M. Morozov, “Machining-based coverage path planning for automated structural inspection,” *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 1, pp. 202–213, Jan. 2018.
- [2] H. Sohn, J. A. Czarnecki, and C. R. Farrar, “Structural health monitoring using statistical process control,” *J. Struct. Eng.*, vol. 126, no. 11, pp. 1356–1363, Nov. 2000.
- [3] A. Entezami, H. Shariyatmadar, and S. Mariani, “Fast unsupervised learning methods for structural health monitoring with large vibration data from dense sensor networks,” *Struct. Health Monit.*, vol. 19, no. 6, 2019, Art. no. 1475921719894186.
- [4] E. Peter Carden and J. M. W. Brownjohn, “ARMA modelled time-series classification for structural health monitoring of civil infrastructure,” *Mech. Syst. Signal Process.*, vol. 22, no. 2, pp. 295–314, Feb. 2008.
- [5] N. G. Pnevmatikos, B. Blachowski, G. D. Hatzigeorgiou, and A. Swiercz, “Wavelet analysis based damage localization in steel frames with bolted connections,” *Smart Struct. Syst.*, vol. 18, no. 6, pp. 1189–1202, Dec. 2016.
- [6] L. Zeng, Y. Liu, G. Zhang, L. Tang, Z. Jiang, and Z. Liu, “Analysis of structural responses of bridges based on long-term structural health monitoring,” *Mech. Adv. Mater. Struc.*, vol. 25, no. 1, pp. 79–86, 2018.
- [7] S.-L. Ma, S.-F. Jiang, C. Wu, and S.-Y. Wu, “Identification of sudden stiffness change in the acceleration response of a nonlinear hysteretic structure,” *Shock Vib.*, vol. 2020, pp. 1–20, Jan. 2020.
- [8] E. J. OBrien, A. Malekjafarian, and A. González, “Application of empirical mode decomposition to drive-by bridge damage detection,” *Eur. J. Mech. A, Solids*, vol. 61, pp. 151–163, Jan. 2017.
- [9] A. Sadhu, “An integrated multivariate empirical mode decomposition method towards modal identification of structures,” *J. Vib. Control*, vol. 23, no. 17, pp. 2727–2741, Oct. 2017.
- [10] F. Xiao, G. S. Chen, J. L. Hulsey, J. D. Dolan, and Y. Dong, “Ambient loading and modal parameters for the chulitna river bridge,” *Adv. Struct. Eng.*, vol. 19, no. 4, pp. 660–670, Apr. 2016.
- [11] Z. Tang, Z. Chen, Y. Bao, and H. Li, “Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring,” *Struct. Control Health Monitor.*, vol. 26, no. 1, p. e2296, Jan. 2019.
- [12] S. F. Masri, M. Nakamura, A. G. Chassiakos, and T. K. Caughey, “Neural network approach to detection of changes in structural parameters,” *J. Eng. Mech.*, vol. 122, no. 4, pp. 350–360, Apr. 1996.
- [13] R.-Y. Yang and R. Rai, “Machine auscultation: Enabling machine diagnostics using convolutional neural networks and large-scale machine audio data,” *Adv. Manuf.*, vol. 7, no. 2, pp. 174–187, Jun. 2019.
- [14] C. Chatfield, *The Analysis of Time Series: An Introduction With R*. Boca Raton, FL, USA: CRC Press, 2003.

- [15] P. Virtanen *et al.*, “Scipy 1.0: Fundamental algorithms for scientific computing in Python,” *Nature Methods*, vol. 17, pp. 261–272, Mar. 2020.
- [16] D. Sundararajan, *Discrete Wavelet Transform: A Signal Processing Approach*. Hoboken, NJ, USA: Wiley, 2016.
- [17] N. E. Huang *et al.*, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
- [18] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, “1D convolutional neural networks and applications: A survey,” 2019, *arXiv:1905.03554*. [Online]. Available: <http://arxiv.org/abs/1905.03554>
- [19] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [20] A. Paszke *et al.*, “Pytorch: An imperative style, high-performance deep learning library,” in *Proc. NeurIPS*, 2019, pp. 8024–8035.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [22] S. Jialin Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [23] J. Howard and S. Gugger, “Fastai: A layered API for deep learning,” *Information*, vol. 11, no. 2, p. 108, Feb. 2020.
- [24] E. Figueiredo, G. Park, J. Figueiras, C. Farrar, and K. Worden, “Structural health monitoring algorithm comparisons using standard data sets,” Los Alamos National Lab., Los Alamos, NM, USA, Tech. Rep. la-14393, 2009.
- [25] K. M. Rashid and J. Louis, “Times-series data augmentation and deep learning for construction equipment activity recognition,” *Adv. Eng. Informat.*, vol. 42, Oct. 2019, Art. no. 100944.
- [26] P. Prasanna *et al.*, “Automated crack detection on concrete bridges,” *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 591–599, Apr. 2016.
- [27] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, *arXiv:1412.6980*. [Online]. Available: <http://arxiv.org/abs/1412.6980>
- [28] Z. Li, C. Lam, J. Yao, and Q. Yao, “On testing for high-dimensional white noise,” *Ann. Statist.*, vol. 47, no. 6, pp. 3382–3412, Dec. 2019.
- [29] B. A. de Castro, F. G. Baptista, and F. Ciampa, “New signal processing approach for structural health monitoring in noisy environments based on impedance measurements,” *Measurement*, vol. 137, pp. 155–167, Apr. 2019.
- [30] R. Brincker and P. Andersen, “ARMA models in modal space,” in *Proc. IMAC*, 1999, pp. 330–334.
- [31] A. Ibrahim, A. Eltawil, Y. Na, and S. El-Tawil, “A machine learning approach for structural health monitoring using noisy data sets,” *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 2, pp. 900–908, Apr. 2020.
- [32] D. Simulia, “Abaqus theory guide,” Dassault Systèmes, Vélizy-Villacoublay, France, Tech. Rep. v6.14, 2016.
- [33] H. Tran-Ngoc *et al.*, “An efficient approach to model updating for a multispan railway bridge using orthogonal diagonalization combined with improved particle swarm optimization,” *J. Sound Vib.*, vol. 476, Jun. 2020, Art. no. 115315.
- [34] V. Ho, T. Hoang, G. De Roeck, T. Bui, and M. Wahab, “Effects of measuring techniques on the accuracy of estimating cable tension in a cable-stay bridge,” in *Proc. DAMAS*, 2020, pp. 433–445.
- [35] L. R. F. D. Aashto, “Bridge design specifications,” Amer. Assoc. State Highway Transp. Officials, Washington, DC, USA, Tech. Rep. 9, 2020.
- [36] J. Maeck, B. Peeters, and G. De Roeck, “Damage identification on the Z24 bridge using vibration monitoring,” *Smart Mater. Struct.*, vol. 10, no. 3, p. 512, 2001.
- [37] L. van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.



Hung V. Dang received the M.Sc. and Ph.D. degrees in structural dynamics from the University of Lyon, Lyon, France, in 2009 and 2013, respectively.

He is a Postdoctoral Fellow with the Faculty of Science and Technology, Middlesex University, London, U.K. His research interests include structural dynamic, numerical simulation, structural health monitoring, data analysis, machine learning, and digital twin. He has published his research works internationally in France, Italy, USA, Hong Kong, and Morocco.



Hoa Tran-Ngoc is currently pursuing the Ph.D. degree with the Department of Electrical Energy, Metals, Mechanical Constructions, and Systems, Ghent University, Ghent, Belgium.

His research interests include structural optimization, machine learning, structural identification, bridge structure engineering, and finite element method. He has published papers in a number of high ranking journals such as *Engineering Structure*, *Journal of Sound and Vibration*, and *Sensors*.



Tung V. Nguyen received the M.S. degree in mechanical engineering with the University of Liège, Liège, Belgium, in 2001, and the Ph.D. degree in material and structure with École des ponts ParisTech, Paris, France, in 2004.

He is an expert in the fields of bridge engineering, design and construction, mechanical engineering, materials and structures. He has had 15 years of working experience in France, with a deep expertise in modeling and simulation, and computational mechanics. Currently, he is the Leader of Modeling and Simulation Team with Schlumberger, Clamart, France.



T. Bui-Tien received the Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia, in 2007.

He then worked as a Post-doctoral Fellow with the Université de Liège, Liège, Belgium and KU Leuven, Leuven, Belgium. He is Vice Dean of the Faculty of Civil Engineering, University of Transport and Communications, Hanoi, Vietnam. His research interests include structural health monitoring, digital twin, intelligent transportation, machine learning, and structural damage diagnosis of bridge.



Guido De Roeck is Emeritus Professor of Faculty of Engineering Science, KU Leuven, Leuven, Belgium, where he is also the Head of the Structural Mechanics Section. He leads multiple research activities ranging a spectrum of domains, including static and dynamic analysis of mechanical structures, damage detection by vibration monitoring, dynamic system identification, soil-structure interaction, nonlinear constitutive soil models, fracture mechanics of composite materials, fuzzy finite element method, and so on.



Huan X. Nguyen (Senior Member, IEEE) received the B.Sc. degree from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2000, and the Ph.D. degree from the University of New South Wales, Sydney, NSW, Australia, in 2007.

He is currently a Professor of digital communication engineering with the Faculty of Science and Technology, Middlesex University, London, U.K., where he is also the Director of the London Digital Twin Research Center and Head of the 5G & IoT Research Group. He leads research activities in 5G systems, machine-type communication, digital transformation, and machine learning within his university with focus on applications in digital twins, industry 4.0, and critical applications (disasters, smart manufacturing, intelligent transportation, e-health). He has been managing many council/industry funded projects and publishing more than 100 peer-reviewed research papers.

Dr. Nguyen has served as chairs for many international conferences such as the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2020 (PIMRC'20), International Conference on Telecommunications (ICT'20), ICT'19, and the EAI International Conference on Forthcoming Networks and Sustainability in the IoT Era (FoNeS-IoT'20).