

RESEARCH ARTICLE

Open Access



# Fully convolutional networks for structural health monitoring through multivariate time series classification

Luca Rosafalco<sup>1</sup>, Andrea Manzoni<sup>2</sup>, Stefano Mariani<sup>1</sup> and Alberto Corigliano<sup>1\*</sup> 

\*Correspondence:

[alberto.corigliano@polimi.it](mailto:alberto.corigliano@polimi.it)

<sup>1</sup> Dipartimento di Ingegneria  
Civile e Ambientale, Politecnico  
di Milano, Piazza L. da Vinci 32,  
20133 Milano, Italy

Full list of author information is  
available at the end of the article

## Abstract

We propose a novel approach to structural health monitoring (SHM), aiming at the automatic identification of damage-sensitive features from data acquired through pervasive sensor systems. Damage detection and localization are formulated as classification problems, and tackled through fully convolutional networks (FCNs). A supervised training of the proposed network architecture is performed on data extracted from numerical simulations of a physics-based model (playing the role of digital twin of the structure to be monitored) accounting for different damage scenarios. By relying on this simplified model of the structure, several load conditions are considered during the training phase of the FCN, whose architecture has been designed to deal with time series of different length. The training of the neural network is done before the monitoring system starts operating, thus enabling a real time damage classification. The numerical performances of the proposed strategy are assessed on a numerical benchmark case consisting of an eight-story shear building subjected to two load types, one of which modeling random vibrations due to low-energy seismicity. Measurement noise has been added to the responses of the structure to mimic the outputs of a real monitoring system. Extremely good classification capacities are shown: among the nine possible alternatives (represented by the healthy state and by a damage at any floor), damage is correctly classified in up to 95% of cases, thus showing the strong potential of the proposed approach in view of the application to real-life cases.

**Keywords:** Structural health monitoring, Fully convolutional networks, Damage localization, Time series analysis, Deep learning

## Introduction

Collapses of civil infrastructures strike public opinion more and more often. They are generally due to either structural deterioration or modified working conditions with respect to the design ones. The main challenge of structural health monitoring (SHM) is to increase the safety level of ageing structures by detecting, locating and quantifying the presence and the development of damages, possibly in real-time [1]. However, visual inspections—whose frequencies are usually determined by the importance and the age of the structure—are still the workhorse in this field, even if they are rarely able to pro-

vide a quantitative estimate of structural damages. Therefore, it is evident why recent advances in sensing technologies and signal processing, coupled to the increased availability of computing power, are creating huge expectations in the development of robust and continuous SHM systems [2].

SHM applications are often treated as *classification problems* [3] aiming (i) to distinguish the damage state of a structure from the undamaged state, starting from a set of available recordings of a monitoring sensor system, and (ii) to locate and quantify the current damage. In this framework, we have adopted the so-called simulation-based classification (SBC) approach [4], and we have exploited deep learning (DL) techniques for the sake of automatic classification. In our procedure, data are displacement and/or acceleration recordings of the structural response, and the classification task consists of recognizing which structural state, among a discrete set, could have most probably produced them. These structural states, characterized by the presence of damage in different positions and of different magnitudes, suitably represent different damage scenarios.

To highlight the distinctive components of the SBC approach, we recall the general paradigm for a SHM system, according to [3]. A SHM system consists of four sequential procedures: (i) operational evaluation, (ii) data acquisition, (iii) features extraction and (iv) statistical inference. Operational evaluation defines what the object of the monitoring is and what the most probable damage scenarios are; data acquisition deals instead with the implementation of the sensing system; features extraction specifies how to exploit the acquired signals to derive features, that is, a reduced representation of the initial data, yet containing all their relevant information—for the case at hand, the onset and propagation of damage in the structure; statistical inference finally sets the criteria under which the classification task is performed.

Focusing on stages (ii) and (iii), the vibration-based approach is nowadays the most common procedure in civil SHM. Its popularity is mainly due to the effective idea that the ongoing damage alters the structure vibration response [5] and, consequently, the associated modal information. By looking at the displacement and/or the acceleration time recordings acquired at a certain set of points of a building, the vibration-based approach enables the analysis of both global and local structural behaviors. The technology required to build this type of sensor system is mature and can be exploited on massive scale [6]. In most of the cases, features extraction relies on determining the system eigenfrequencies and the modal shapes. On the other hand, it might be necessary to employ more involved outcomes to distinguish between the effect of modified loading conditions and the true effect of damage [7], for instance by constructing parametric time series models [8]. By employing DL, we aim at dealing with these aspects automatically.

Two competing approaches are employed in literature to deal with stage (iv), the a) model-based and the b) data-based approach, both introducing a sort of offline–online decomposition. By this expression, we mean the possibility to split the procedure into two phases: first, the offline phase is performed before the structure starts operating; then, the online phase is carried out during its normal operations.

The model-based approach builds a physics-based model, initially calibrated to simulate the structural response. The model is updated whenever new observations become available and, accordingly, damage is detected and located. Data assimilation techniques such as Kalman filters have been employed to efficiently deal with model updating [9].

Model-based approaches are typically ill-conditioned, and many uncertainties related to the proper tuning of model parameters may prevent a correct damage estimation.

Hence, data-based approaches are becoming more and more popular; they exploit a collection of structural responses and, either assess any deviation between real and simulated data, or assign to the measured data the relevant class label. The dataset construction can be done either experimentally [10] or numerically; however, the latter option is usually preferred, due to the frequent difficulties in reproducing the effects of damage in real-scale civil structures properly. To reduce the computational burden associated with the dataset construction, simplified models (e.g. mass-spring models for the dynamics of tall and slender buildings)—still able to catch the correct structural response—are preferred with respect to more expensive high-fidelity simulations, involving, e.g., the discretization of both structural and non-structural elements. By adopting the SBC method, we rely on a data-driven approach based on synthetic experiments.

Once a dataset of possible damage scenarios has been constructed, machine learning (ML) proved to be suitable to perform the classification task [6]. The training of the ML classifier could be:

- supervised, when a label corresponding to one of the possible outputs of the classification task is associated to each structural response;
- unsupervised [11], when no labelling is available;
- semi-supervised [12], when the training data only refer to a reference condition.

In the SBC framework, a semi-supervised approach was recently explored, e.g., in [13], leading to great computational savings and robust results when treating the anomaly detection task. In spite of their good performances, standard ML techniques based, e.g., on statistical distributions of the damage classes (as in the so-called decision boundary methods), as well as kernel-based methods (e.g. support vector machines), still rely on heavy data preprocessing, required to compute problem-specific sets of engineered features [14]. These features can be statistics of the signal, modal properties of the structure, or even more involved measures exploiting different types of signal transformation (e.g. Power Spectral Density and autocorrelation functions, to mention a few) [6]. Some relevant drawbacks arise, since:

- pre-computed engineered features are not well suited for non-standard problems, for which setting damage classification criteria can be anything but trivial;
- there is no way to assess the optimality of the employed features;
- a computationally expensive pre-processing of a huge amount of data is usually required.

For these reasons, we rely on deep learning techniques, which allow both data dimensionality reduction and hierarchical pattern recognition at the same time [15,16]. DL techniques allow us to:

- deal with non-standard problems, especially when different information sources have to be managed (as long as they are in the form of time series);
- detect a set of features, optimized with respect to the classification task, through the training of an artificial neural network.

Despite these advantages, the use of DL for the sake of SHM has been quite limited so far [17, 18]. We have therefore decided to employ Fully Convolutional Networks (FCNs) [19], a particular Neural Network (NN) architecture, to deal with the Multivariate Time Series (MTS) produced by monitoring sensor systems. To face different information sources, we have applied separate convolutional branches and, at a second stage, performed the data fusion of the extracted information.

### SHM methodology

We introduce in this section a detailed explanation of the proposed strategy to deal with the SHM problem exploiting a SBC approach. We provide a simplified physics-based model of the structure employing  $M$  degrees of freedom (dofs), assuming to record time-dependent signals through a monitoring system employing  $N_0 \leq M$  sensors. Our aim is first to train, and then to use, two classifiers  $\mathcal{G}_d$  and  $\mathcal{G}_l$  for the sake of damage detection and localization, respectively, where

$$\mathcal{G}_d : \mathbb{R}^{N_0 \times L_0} \rightarrow \{0, 1\}, \quad \mathcal{G}_l : \mathbb{R}^{N_0 \times L_0} \rightarrow \{0, 1, \dots, G\}.$$

In the former case, labels 0 and 1 denote absence or presence of damage, respectively; in the latter,  $G > 1$  is a priori fixed and denotes the range of possible damage locations—also in this case, the undamaged state is denoted by 0. We have decided to include the undamaged state among the possible outputs of  $\mathcal{G}_l$  not just to confirm the outcome of  $\mathcal{G}_d$ , but also to observe which damage scenarios, identified by their locations, are more often misclassified with the undamaged state.

The training of  $\mathcal{G}_d$  and  $\mathcal{G}_l$  is performed using the two datasets  $\mathbb{D}_{train}^d$  and  $\mathbb{D}_{train}^l$ , respectively. Each of these two datasets (for simplicity we only consider the formation of  $\mathbb{D}_{train}^l$ , being the process substantially equivalent for  $\mathbb{D}_{train}^d$ ) collects  $V^{train}$  structural responses,

$$\mathbb{D}_{train}^l = \{\mathbb{U}_1, \dots, \mathbb{U}_{V^{train}}\},$$

under prescribed damage scenarios and loading conditions. We denote by  $\mathbb{U}_i \in \mathbb{R}^{N_0 \times L_0}$ ,  $i = 1, \dots, V^{train}$ , a collection of  $N_0$  sensor recordings of displacement and/or acceleration time series of length  $L_0$ , such that

$$\mathbb{U}_i = [\mathbf{u}_1(\mathbf{d}_i, \mathbf{l}_i) \mid \dots \mid \mathbf{u}_{N_0}(\mathbf{d}_i, \mathbf{l}_i)], \quad i = 1, \dots, V^{train}; \quad (1)$$

the time series  $\mathbf{u}_n(\mathbf{d}_i, \mathbf{l}_i)$  recorded by the  $n$ -th sensor depends on the damage scenario  $\mathbf{d}_i$  and the loading condition  $\mathbf{l}_i$ , and can be seen as the sampling of a time-dependent signal  $\mathbf{u}_n(\mathbf{d}_i, \mathbf{l}_i)$ . We assume to deal with recordings acquired at a set of  $L_0$  time instants uniformly distributed over the time interval of interest  $I$ . The damage scenario  $\mathbf{d}_i : \mathcal{P}_d \rightarrow \mathbb{R}^M$  is prescribed at each structural element<sup>1</sup> and depends on a set of parameters  $\boldsymbol{\eta}_d \in \mathcal{P}_d \subset \mathbb{R}^D$ ; the loading condition  $\mathbf{l}_i : I \times \mathcal{P}_l \rightarrow \mathbb{R}^M$ , defined over the time interval  $I$ , is prescribed at each element, too, and depends on a set of parameters  $\boldsymbol{\eta}_l \in \mathcal{P}_l \subset \mathbb{R}^L$ . Here, we denote by  $\mathcal{P}_d$  and  $\mathcal{P}_l$  two sets of parameters, yielding the two sets  $\mathcal{C}_d$  and  $\mathcal{C}_l$  of admissible damage and loading scenarios, respectively, obtained when sampling  $\boldsymbol{\eta}_d \in \mathcal{P}_d$  and  $\boldsymbol{\eta}_l \in \mathcal{P}_l$ . During the training procedure, the performances of  $\mathcal{G}_d$  and  $\mathcal{G}_l$  are tracked by looking at their classification capabilities on two datasets  $\mathbb{D}_{val}^d$  and  $\mathbb{D}_{val}^l$ , each one collecting  $V^{val}$  structural responses  $\mathbb{U}_i$  (defined as in Eq. (1)),  $i = 1, \dots, V^{val}$ .

<sup>1</sup> For simplicity, the number  $E$  of elements coincides with the number  $M$  of degrees of freedom; however, the generalization to the case in which  $E \neq M$  is straightforward.

According to the SBC approach, the datasets  $\mathbb{D}_{train}^d, \mathbb{D}_{train}^l, \mathbb{D}_{val}^d$  and  $\mathbb{D}_{val}^l$  are constructed by exploiting a simplified physics-based model of the structure. For any damage scenario  $\mathbf{d} \in \mathcal{C}_d$  and loading conditions  $\mathbf{l} \in \mathcal{C}_l$  received as inputs, this numerical model—playing the role of digital twin of the structure to be monitored—returns a recorded displacement and/or acceleration time series  $\mathbf{r}_n(\mathbf{d}, \mathbf{l})$ . Since these latter are deterministic, to make our data more conformal to real measurements  $\mathbf{u}_n(\mathbf{d}, \mathbf{l})$ , we assume that each  $\mathbf{r}_n(\mathbf{d}, \mathbf{l})$  is affected by an additive measurement noise  $\boldsymbol{\epsilon}_n \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$ , so that

$$\mathbf{u}_n = \mathbf{r}_n(\mathbf{d}, \mathbf{l}) + \boldsymbol{\epsilon}_n, \quad n = 1, \dots, N_0. \quad (2)$$

Here we consider each  $\boldsymbol{\epsilon}_n$  normally distributed, with zero mean and covariance matrix  $\boldsymbol{\Sigma}_\epsilon \in \mathbb{R}^{N_0 \times N_0}$ , as related to a real monitoring system [20]. Regarding the auto-correlation of the records ( $j = 1, \dots, L_0$ ) of each sensor ( $n = 1, \dots, N_0$ ) in time, we assume them to be independent and identically distributed.

The background model providing  $\mathbf{r}_n(\mathbf{d}, \mathbf{l})$  is here thought as being already tuned to accurately match the structural response in the undamaged case. Moving away from the baseline due to damage inception, with the adopted supervised strategy we therefore assume the possible damage scenarios to belong to a limited set, and for each of them relevant numerical analyses are exploited to mimic the real structural response, as affected by all the possible uncertainty sources. It should be also added that Eq. (2) accounts for the noise in the structural response induced by sensor measurements only. Since damage is a smeared measure of different phenomena occurring at the local scale (including or accompanied by, e.g. cracking and plasticity), it stands as a variable giving a measure of the unresolved dofs in a Mori–Zwanzig formalism, see [21]. In a state-space formulation like the one adopted for Kalman filtering [22], a further source of noise can be added through the state or model error, which accounts for the uncertainties linked to the unresolved dynamics of the system. An issue may thus arise in discerning the two noise sources linked to the model inaccuracy on one side, and to the sensor output and operational conditions on the other side. This discussion is indeed beyond the scope of this work, and interested readers may find relevant information in, e.g. [23, 24].

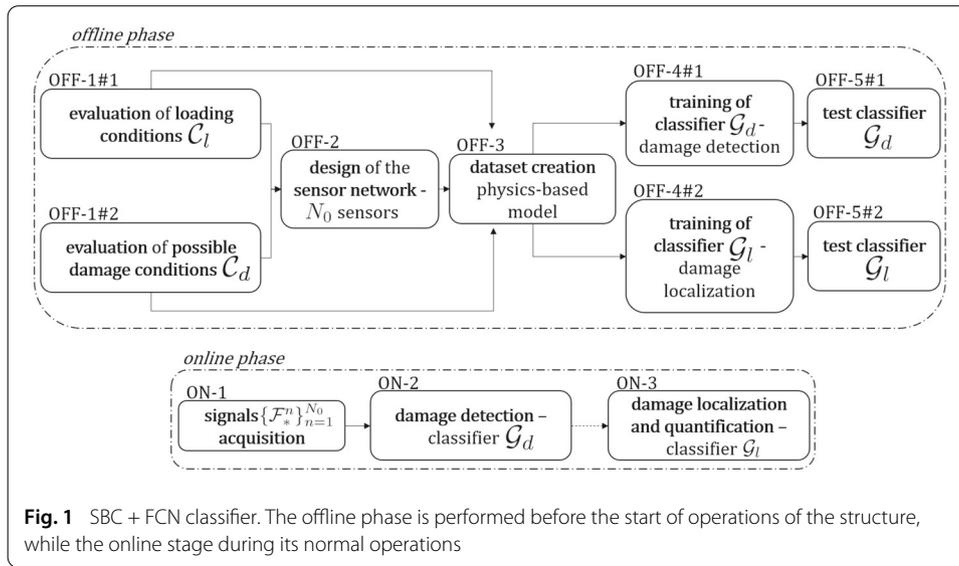
The classifiers  $\mathcal{G}_d$  and  $\mathcal{G}_l$  are based on a fully convolutional neural network architecture (that will be detailed in the following section). The training of the network is supervised, and performed by feeding the FCN with multivariate time series  $\{\mathcal{F}_0^n\}_{n=1}^{N_0}$  and associated labels (0 or 1 for  $\mathcal{G}_d, g \in \{0, 1, \dots, G\}$  for  $\mathcal{G}_l$ ). In this respect, hereon each multivariate time series  $\{\mathcal{F}_0^n\}_{n=1}^{N_0}$  is referred to as an instance. In general,  $\{\mathcal{F}_0^n\}_{n=1}^{N_0} = \mathbb{U}_i$ ; however, a single instance might be made up to  $W$  multivariate time series  $\mathbb{U}_{iw}, w = 1, 2, \dots, W$  of different lengths  $L_0^w$  to deal with the case of sensors recording time series of different length. Each component  $\mathcal{F}_0^n = \mathbf{u}_n$  plays the role of input channel for the NN.

The testing of the NN is done on instances  $\{\mathcal{F}_*^n\}_{n=1}^{N_0} = \mathbb{U}_i^*$ , obtained through the numerical model as structural response

$$\mathbb{U}_i^* = [\mathbf{u}_1(\mathbf{d}_i, \mathbf{l}_i^*) \mid \dots \mid \mathbf{u}_{N_0}(\mathbf{d}_i, \mathbf{l}_i^*)], \quad i = 1, \dots, V^{test}$$

to loading conditions  $\mathbf{l}_i^* \in \mathcal{C}_l, i = 1, \dots, V^{test}$ , unseen (that is, associated to testing values  $\eta_l$  from  $\mathcal{P}_l$  not sampled) when building the datasets  $\mathbb{D}_{train}^d, \mathbb{D}_{train}^l, \mathbb{D}_{val}^d$  and  $\mathbb{D}_{val}^l$ . All these instances are collected into two datasets  $\mathbb{D}_{test}^d$  and  $\mathbb{D}_{test}^l$ .

The testing is done by verifying the correct identification of the class ( $\{0, 1\}$  for  $\mathcal{G}_d, \{0, 1, \dots, G\}$  for  $\mathcal{G}_l$ ) associated with the simulated signals. In concrete terms, a probability



is estimated for each possible class, thus yielding the confidence level that the given class is assigned to the data, and the class with highest confidence is compared with the one associated to the simulated signal. No k-fold cross validation is used.

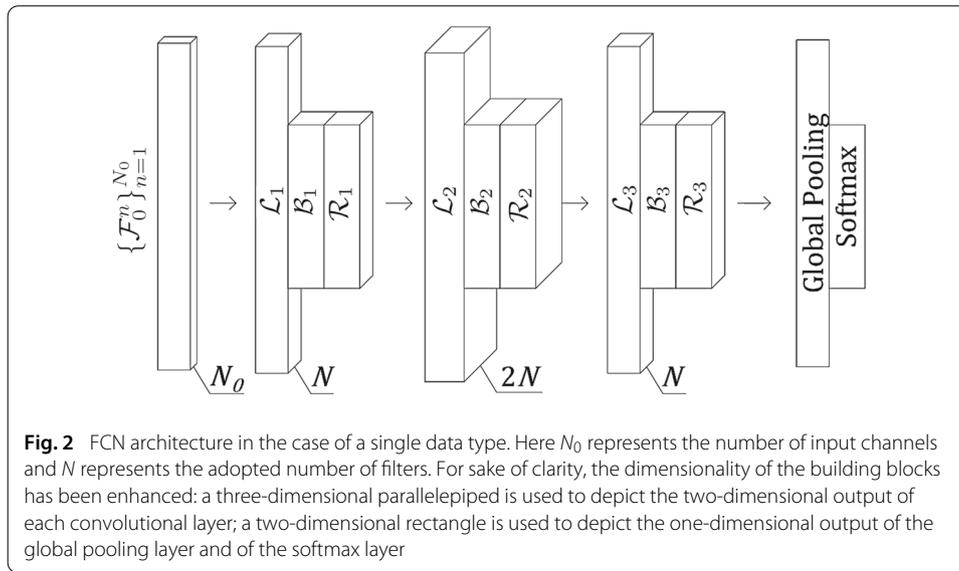
Once tested,  $\mathcal{G}_d$  and  $\mathcal{G}_l$  can make a prediction once a new signal  $\{\mathcal{F}_*^n\}_{n=1}^{N_0} = \mathbb{U}^*$  is experimentally acquired from the real sensor network used to monitor the structure.

Let us now recap the procedure steps exploiting the schematic representation reported in Fig. 1. For the sake of convenience, we can split our procedure into:

- an offline phase, where, as first step, the loading conditions  $\mathcal{C}_l$  (OFF-1#1) and the most probable damage scenarios  $\mathcal{C}_d$  are evaluated (OFF-1#2). Accordingly, a sensor network with  $N_0$  sensors is designed (OFF-2). The datasets  $\mathbb{D}_{train}^d$ ,  $\mathbb{D}_{train}^l$ ,  $\mathbb{D}_{val}^d$ ,  $\mathbb{D}_{val}^l$ ,  $\mathbb{D}_{test}^d$  and  $\mathbb{D}_{test}^l$  are then constructed (OFF-3) by exploiting the physics-based *digital twin* of the structure. The classifiers  $\mathcal{G}_d$  (OFF-4#1) and  $\mathcal{G}_l$  (OFF-4#2) are therefore trained by using  $\mathbb{D}_{train}^d$  and  $\mathbb{D}_{train}^l$  and performing the validation using  $\mathbb{D}_{val}^d$  and  $\mathbb{D}_{val}^l$ . Finally, the classification capacity of  $\mathcal{G}_d$  and  $\mathcal{G}_l$  is assessed by using numerically simulated signals  $\{\mathcal{F}_*^n\}_{n=1}^{N_0} = \mathbb{U}^*$  belonging to  $\mathbb{D}_{test}^d$  and  $\mathbb{D}_{test}^l$ , respectively (OFF-5#1 and OFF-5#2);
- an online phase, in which for any new signal  $\{\mathcal{F}_*^n\}_{n=1}^{N_0} = \mathbb{U}^*$  acquired by the real monitoring system and provided to the classifiers (ON-1), damage detection (ON-2) is performed through  $\mathcal{G}_d$ , and damage localization is performed through  $\mathcal{G}_l$  (ON-3).

In lack of recordings coming from a real monitoring system, and having assumed the experimental signals  $\mathbb{U}^*$  equal to the noise-corrupted output of the numerical model, steps OFF-5#1 and OFF-5#2 of the offline phase indeed coincide with steps ON-2 and ON-3 of the online procedure.<sup>2</sup> We highlight that only those damage scenarios  $\mathbf{d} \in \mathcal{C}_d$  that have been numerically simulated in the offline phase can be classified during the online phase. Moreover, damage is considered temporary frozen within a fixed observation interval,

<sup>2</sup> For this reason, the acquired signals are denoted with the same notation  $\mathbb{U}^*$  employed for the recordings previously used to test the FCN to highlight that, for the time being, the experimental signals are taken as realizations of the noise-corrupted outputs of the numerical model.



enabling to treat the structure as linear [2]. To model the effect of damage, we consider the stiffness degradation of each structural member; this assumption is acceptable if the rate of the evolving damage is sufficiently small with respect to the observation interval [25].

It is not possible to identify from the beginning the most suitable number of instances  $V^{train}$  to be used to train the network. The easiest procedure (even if time-consuming) would be to assess the performances of  $\mathcal{G}_d$  and  $\mathcal{G}_l$  for different sizes  $V^{train}$ , aiming at finding a trade-off between the computational burden required to construct the dataset and train the NN, and the classification capabilities. Beyond a certain critical size, massive dataset enlargements might lead to small improvements in the NN performance, as shown in our numerical results.

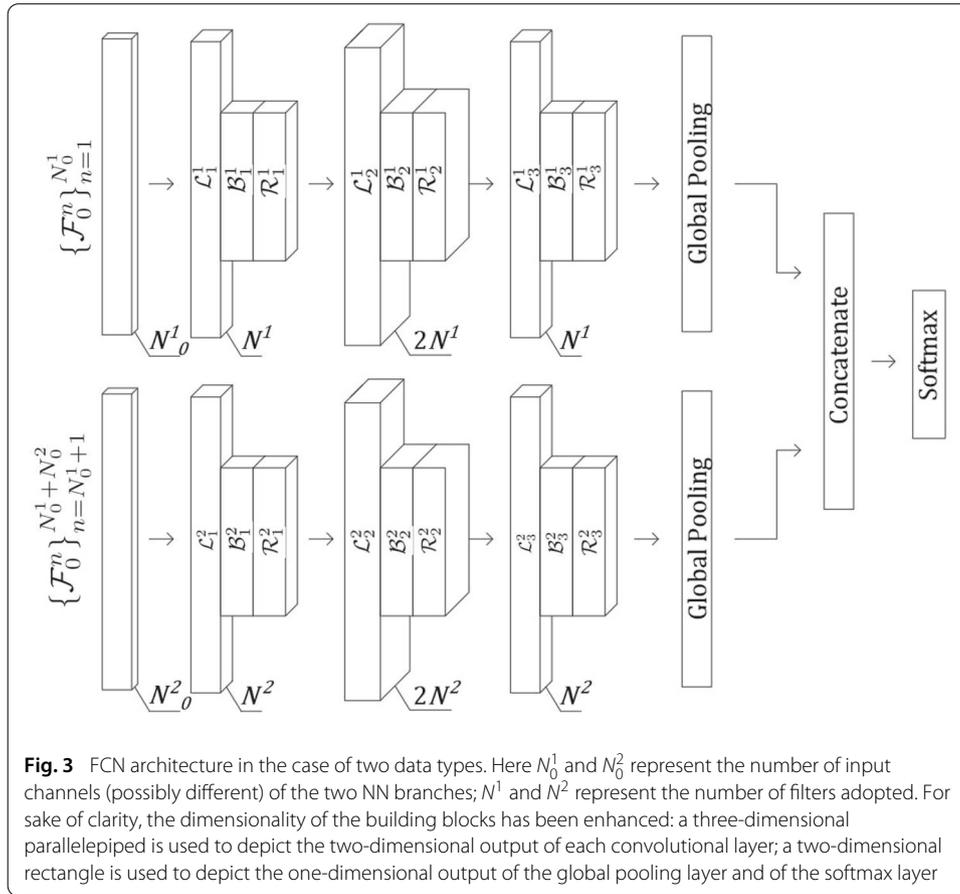
Finally, concerning the setting of the loading conditions  $\mathcal{C}_l$ , in this work we have (i) identified a set of possible loading scenarios that can significantly affect the response of the structure; (ii) subdivided this set into a certain number of subsets, representative of different possible dynamic effects of the applied load; (iii) sampled each subset almost the same number of times.

## Fully convolutional networks

### Neural network architecture

We now describe the FCN architecture employed for the sake of classification. As discussed in the previous section,  $\{\mathcal{F}_0^n\}_{n=1}^{N_0}$  are the inputs adopted during the training phase (for which we know the instance label associated), while  $\{\mathcal{F}_*^n\}_{n=1}^{N_0}$  are the inputs that we require the FCN to classify.

We have adopted a FCN stacking three convolutional layers  $\mathcal{L}_i$ ,  $i = \{1, 2, 3\}$ , with different filter sizes  $h_i$ , followed by a global pooling layer and a softmax classifier (the choice of the NN hyperparameters will be discussed in the following). Each convolutional layer  $\mathcal{L}_i$  has been used together with a Batch-Normalization (BN) layer  $\mathcal{B}_i$  and a Rectified



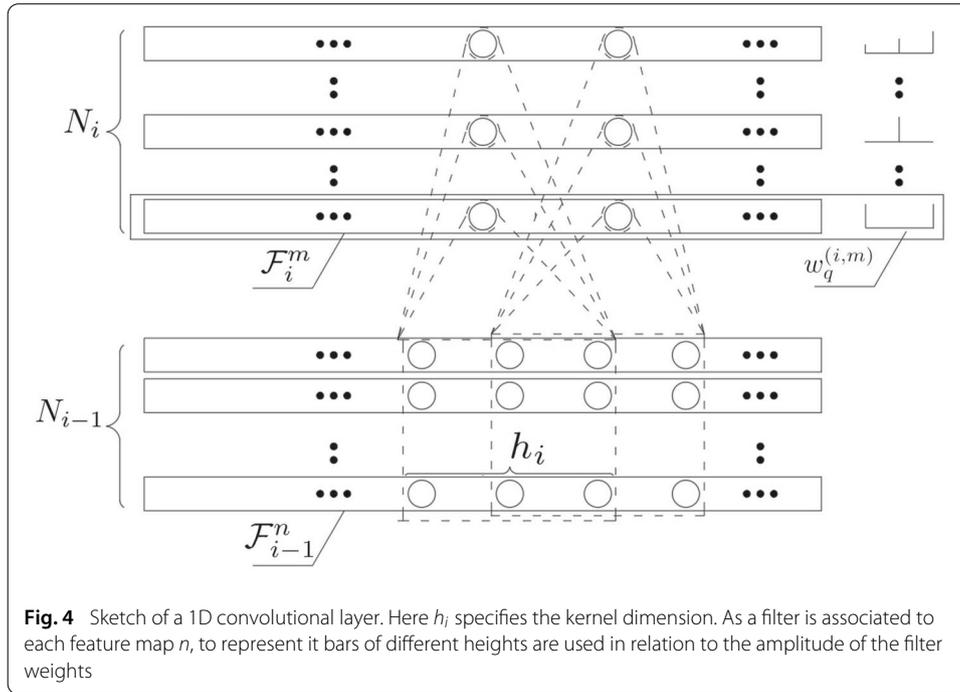
Linear Unit (ReLU) activation layer  $\mathcal{R}_i$  [14, 19], see Fig. 2. When the input signals are made up by  $W$  multivariate time series with different length:

$$\{\mathcal{F}_0^n\}_{n=1}^{N_0^1+\dots+N_0^j+\dots+N_0^W} = \begin{cases} \{\mathcal{F}_0^n\}_{n=1}^{N_0^1} = \mathbb{U}_{i1} \in \mathbb{R}^{N_0^1 \times L_0^1} \\ \{\mathcal{F}_0^n\}_{n=N_0^1+1}^{N_0^1+N_0^2} = \mathbb{U}_{i2} \in \mathbb{R}^{N_0^2 \times L_0^2} \\ \vdots \\ \{\mathcal{F}_0^n\}_{n=N_0^1+\dots+N_0^{W-1}+1}^{N_0^1+\dots+N_0^W} = \mathbb{U}_{iW} \in \mathbb{R}^{N_0^W \times L_0^W} \end{cases},$$

for each one we first adopt the described convolutional architecture separately and then, through a concatenation layer, we perform data fusion on the extracted features. Classification is finally pursued through a softmax layer. The corresponding NN architecture is sketched in Fig. 3 in the case of time series with two different lengths  $L_0^1$  and  $L_0^2$ , but can be easily generalised. Tensorflow [26] has been used for the sake of NN construction.

**Use of convolutional layers**

Let us now show how convolutional layers can be adopted to extract features from multivariate time series.  $\{\mathcal{F}_0^n\}_{n=1}^{N_0}$  are provided to the 1-st convolutional layer  $\mathcal{L}_1$ . The output of  $\mathcal{L}_1$ ,  $\{\mathcal{F}_1^n\}_{n=1}^{N_1}$ , still shaped as time series (of length  $L_1$ ), do not represent displacement and/or acceleration any more. Indeed, they are features extracted from the input channels  $\{\mathcal{F}_0^n\}_{n=1}^{N_0}$ . The following layers operate in the same manner: the outputs  $\{\mathcal{F}_i^n\}_{n=1}^{N_i}$  of the



**Fig. 4** Sketch of a 1D convolutional layer. Here  $h_i$  specifies the kernel dimension. As a filter is associated to each feature map  $n$ , to represent it bars of different heights are used in relation to the amplitude of the filter weights

$(i - 1)$ -th convolutional layer  $\mathcal{L}_{i-1}$  are the inputs of the  $i$ -th convolutional layer  $\mathcal{L}_i$  and become features of higher and higher level.

In concrete terms, the tasks performed by the  $i$ -th convolutional layer  $\mathcal{L}_i$  are: the subdivision of the inputs  $\{\mathcal{F}_{i-1}^n\}_{n=1}^{N_{i-1}}$  into data sequences, whose length  $h_i$  determines the receptive field of  $\mathcal{L}_i$ ; and the multiplication of each data sequence by a set of weights  $w^{(i,m)}$  called filter, where the output  $\mathcal{F}_i^m$  of each filter is called feature map. Mono-dimensional (1D) receptive field must be used in time series analysis, being each channel monodimensional. In Fig. 4 the fundamental architecture of  $\mathcal{L}_i$  is depicted, linking the inputs  $\{\mathcal{F}_{i-1}^n\}_{n=1}^{N_{i-1}}$  and the outputs  $\{\mathcal{F}_i^m\}_{m=1}^{N_i}$  through:

$$z_h^{(i,m)} = \sum_{q=0}^{h_i-1} \sum_{n=1}^{N_{i-1}} w_q^{(i,m)} x_p^{(i-1,n)} + b^{(i,m)} \quad \text{with } p = h + q, \tag{3}$$

where:

- $z_h^{(i,m)}$  is the  $h$ -th entry of  $\mathcal{F}_i^m$ ;
- $b^{(i,m)}$  is the bias of  $\mathcal{F}_i^m$ ;
- $x_p^{(i-1,n)}$  is the  $p$ -th entry of  $\mathcal{F}_{i-1}^n$ ;
- $w_q^{(i-1,n)}$  is the  $q$ -th connection weight of the  $m$ -th filter applied to the  $p$ -th input of  $\mathcal{F}_{i-1}^n$ .

As the goal of stacking several convolutional layers is to provide nonlinear transformations of  $\{\mathcal{F}_0^n\}_{n=1}^{N_0}$ , their overall effect is to make the classes to be recognised linearly separable [27]. In this way, a linear classifier is suitable to carry out the final task. Every nonlinear transformation can be interpreted, as discussed, as an automatic extraction of features.

### Batch Normalization, ReLU activation, global pooling and softmax classifier

The Batch Normalization (BN) layer  $\mathcal{B}_i$  is introduced after each convolutional layer  $\mathcal{L}_i$  to address the issue related to the vanishing/exploding gradients possibly experienced during the training of deep architectures [28]. It relies on normalization and zero-centering of the outputs  $\{\mathcal{F}_i^n\}_{n=1}^{N_i}$  of each layer  $\mathcal{L}_i$ . We express the output of  $\mathcal{B}_i$  as  $\{\mathcal{F}_{\mathcal{B}_i}^n\}_{n=1}^{N_i}$ . For the same reason, the ReLU activation function is preferred instead of saturating ones [29]. The ReLU layer  $\mathcal{R}_i$  transforms  $\{\mathcal{F}_{\mathcal{B}_i}^n\}_{n=1}^{N_i}$ , through

$$\mathcal{F}_{\mathcal{R}_i}^n(u) = \max(0, \mathcal{F}_{\mathcal{B}_i}^n(u)) \quad \text{with } u = 1, \dots, L_i.$$

where:

- $\mathcal{F}_{\mathcal{B}_i}^n(u)$  is the  $u$ -th entry of the  $n$ -th feature map of  $\mathcal{B}_i$ ;
- $\mathcal{F}_{\mathcal{R}_i}^n(u)$  is the  $u$ -th entry of the  $n$ -th feature map of  $\mathcal{R}_i$ .

In the adopted FCN architecture, the features to be used in the classification task are extracted from  $\{\mathcal{F}_0^n\}_{n=1}^{N_0}$  by the blocks  $\{\mathcal{L}_i + \mathcal{B}_i + \mathcal{R}_i\}_{i=1}^3$ . The final number of features equals the number  $N_3$  of filters of the last convolutional layer. By applying next a global average pooling [30], the extracted features  $\{\mathcal{F}_{\mathcal{R}_3}^n\}_{n=1}^{N_3}$  are condensed in a single channel  $\mathbf{b} \in \mathbb{R}^G$ , being  $G$  the total number of classes.

The softmax activation layer finally performs the classification task. First, the channel  $\mathbf{b}$  is mapped onto the target classes, by computing a score  $s_g$

$$s_g(\mathbf{b}) = \boldsymbol{\theta}_g^T \cdot \mathbf{b}, \quad g = 1, \dots, G, \quad (4)$$

for each class  $g$ , where the vector  $\boldsymbol{\theta}_g \in \mathbb{R}^G$  collects the weights related to the  $g$ -th class.

The softmax function is then used to estimate the probability  $p_g \in [0, 1]$  that the input channels belongs to the  $g$ -th class, according to:

$$p_g = \frac{e^{s_g(\mathbf{b})}}{\sum_{j=1}^G e^{s_j(\mathbf{b})}} \quad g = 1, \dots, G. \quad (5)$$

The input channels  $\{\mathcal{F}_0^n\}_{n=1}^{N_0}$  are then assigned to the class with associated label  $g$  featuring the highest estimated probability  $p_g$ , which then represents the estimated confidence level that class  $g$  is assigned to the data.

### Neural Network training

The NN training consists of tuning the weights  $\mathbf{w}^{(i,n)}$  and  $\boldsymbol{\theta}_g$ , respectively appearing in Eqs. (3) and (4) by minimizing a loss function depending on the data. In this respect, the Adam optimization method [31], a widespread stochastic gradient-based optimization method, has been used. For classification purposes, the most commonly adopted loss function is the cross entropy, defined for the classifier  $\mathcal{G}_d$  as:

$$J_d(\mathbf{Y}, \mathbf{p}) = -\frac{1}{V^{train}} \sum_{i=1}^{V^{train}} \sum_{g=1}^G y_i^g \log(p^g), \quad (6)$$

where:

- $g$  is the label of the instance provided to the NN during the training;
- $y_i^g \in \{0, 1\}$  is the confidence that the  $i$ -th instance should be labelled as the  $g$ -th class, with

$$y_i^g = \begin{cases} 1 & \text{if for the } i\text{-th instance the } g\text{-th class is the target class} \\ 0 & \text{otherwise;} \end{cases}$$

- $\mathbf{Y} \in \{0, 1\}^{V^{train}}$  collects all the  $y_i^g$  confidence values;
- $\mathbf{p} \in \mathbb{R}^G$  collects the estimated probabilities  $p^g$ , see Eq. (5).

The loss function  $J_l(\mathbf{Y}, \mathbf{p})$  for the classifier  $\mathcal{G}_l$  is defined analogously.

Regarding the employed datasets:

- $\mathbb{D}_{train}^d$  is used to train the NN by back-propagating the classification error;
- $\mathbb{D}_{val}^d$  is used to possibly interrupt the training in case of overfitting, but not to modify the NN weights;
- $\mathbb{D}_{test}^d$  is used to verify the prediction capabilities of the NN, after the training phase has been performed.

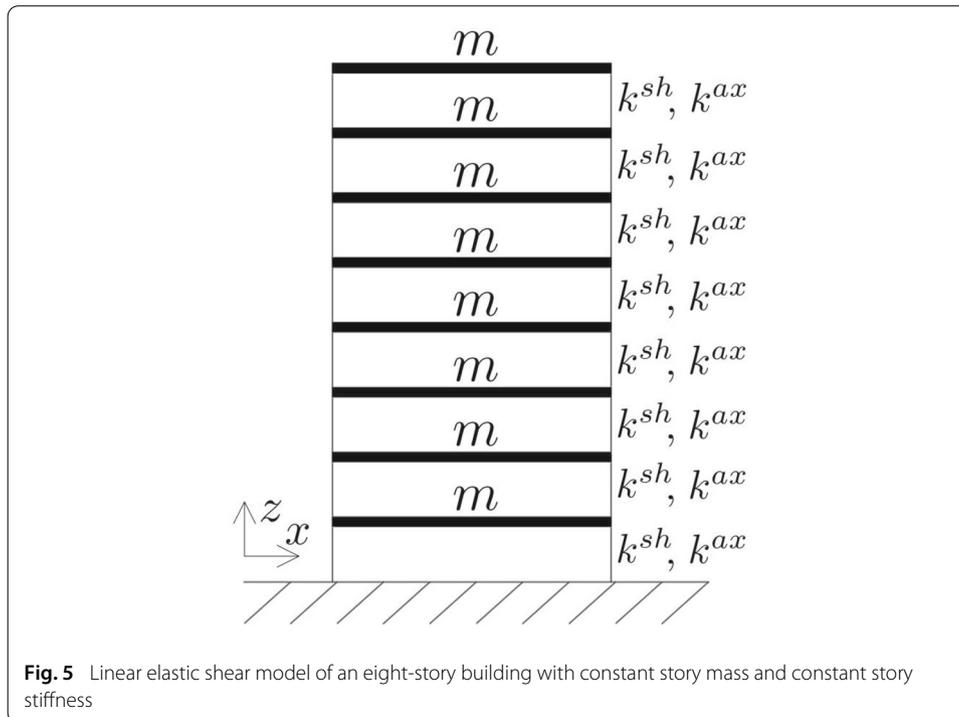
The same splitting applies to the data used for training  $\mathcal{G}_l$ . In order to assess the offline phase of the proposed procedure, we have tested  $\mathcal{G}_d$  and  $\mathcal{G}_l$  on their respective test sets  $\mathbb{D}_{test}^d$  and  $\mathbb{D}_{test}^l$  (steps OFF5#1 and OFF5#2 of Fig. 1). The number of times  $\mathbb{D}_{train}^d$  and  $\mathbb{D}_{train}^l$  are evaluated during the training of  $\mathcal{G}_d$  and  $\mathcal{G}_l$  corresponds to the number of epochs: in this work, we have bounded to 1500 the maximum number of epochs allowed. We have also provided the possibility of an early-stop of the training when, after having performed at least 750 epochs, the validation loss has not decreased three times in a row.

To control the training process, a learning rate  $\xi$  is usually introduced to scale the correction of the NN weights provided by back-propagating the classification error. In our case, the learning rate has been forced to decrease linearly with the number of epochs, moving from  $10^{-3}$  at the beginning of the training till  $10^{-4}$  at its end [32]. After having performed at least 750 epochs, an additional factor  $\zeta = 1/\sqrt[3]{2}$  is used to scale down the learning rate if the loss function  $J(\mathbf{Y}, \mathbf{p})$  is not reduced within the successive 100 epochs, as suggested in [32]. Random subsamples (also called minibatches) of the data points belonging to the training set are employed for the sake of gradient evaluation when running the Adam optimization method [27, 31].

### Hyperparameters setting

The setting of the NN hyperparameters, namely the dimensions of the kernels  $h_i$  and the number of feature maps  $N_i$ , is done according to [14, 32]. In this work, we choose  $h_1 = 8$ ,  $h_2 = 5$ ,  $h_3 = 3$  as kernel dimensions for the three convolutional layers. Since no zero-padding has been employed, the dimension of the time series is progressively reduced passing through the convolutional layer  $\mathcal{L}_i$  from  $L_{i-1}$  to  $L_i = L_{i-1} - h_i + 1$ . Accordingly, considering the parameters and the length of the time series used in this work, the dimension reduction related to a single convolutional layer is on the order of 1%. We have verified that the classification accuracy is barely affected by this reduction and, more in general, by the use of the zero-padding. It is possible to further improve the NN performances by operating a (necessarily problem-dependent) finer tuning of the NN hyperparameters, but only at the cost of a time-consuming repeated evaluation of the NN outcomes.

The number of filters  $N_i$  to be adopted depends on the complexity of the classification task: the more complex the classification, the higher the number of filters needed. However, increasing the number of filters beyond a certain threshold, which depends on the problem complexity and the task to be performed, has no effects on the prediction capabilities of the NN; indeed, the risk would be to increase computational costs, and to overfit



the training dataset. Therefore, it looks convenient to initially employ a small number of filters, and then increase it if the NN performs poorly during the training phase. A possible choice suggested in [14] is to consider  $N_1 = 128$ ,  $N_2 = 256$ , and  $N_3 = 128$  as a suitable choice, independently of the dataset to be analysed. Here we have kept the proportion  $N_1 = N$ ,  $N_2 = 2N$ , and  $N_3 = N$  as filter sequence, and verified that increasing  $N$  beyond  $N = 16$  does not affect the NN performances. To carry out the comparison of FCN architectures with one or two convolutional branches, we have kept  $N = 16$  independently of the classification task.

## Numerical results

### Dataset construction

The proposed methodology is now assessed through the numerical benchmark shown in Fig. 5, and originally proposed in [33]. No real experimental measurements have been allowed for in the analysis; measurement noise has been instead introduced by corrupting the monitored structural response with uncorrelated random signals featuring different Signal to Noise Ratio (SNR) levels, to also assess the effect of sensors accuracy on the capability of the proposed approach. Further details are provided below.

The considered structure is an idealised eight-story shear building model, featuring a constant floor mass of  $m = 625$  t and a constant inter-story stiffness of  $k^{sh} = 10^6$  kN/m. The proposed SHM strategy has been designed to handle signals related to different types of damage-sensitive structural responses characterized by different magnitude and sampling rate. Hence, in the following both the horizontal and the vertical motions of each story are allowed for and recorded. The longitudinal stiffness of the columns has been set to  $k^{ax} = 10^8$  kN/m, and a slenderness (given by the ratio their length and thickness) of 10 has been assumed for the same columns. The numerical model employs  $M = 16$  dofs

(8 in the  $x$  direction and 8 in the  $z$  direction), and  $N_0 = 16$  virtual sensors are used to measure the noise-free displacements  $\mathbf{r}_n$  (collecting both horizontal displacements  $\mathbf{r}_n$ ,  $n = 1, \dots, 8$ , and vertical displacements  $\mathbf{r}_n$ ,  $n = 9, \dots, 16$  the vertical displacements) at all the story levels. Although a non-classical damping was originally proposed in [33], the relevant effect on system identification or model update has shown to be marginal if the structure is continuously excited during the monitoring stage, see e.g. [2,34]. Therefore, in this feasibility study no damping has been taken into account. The dofs are numbered from 1 for the ground floor up to 8 for the eighth floor in both directions.

Due to the building geometry, eight different damage scenarios  $\mathbf{d}(1), \dots, \mathbf{d}(8)$  can be considered, each one characterized by a reduction of 25% of one inter-story stiffness only, that is,

$$\mathbf{d}(g) = \begin{cases} 0.75k_j & \text{if } j = g \text{ or } j = g + 8 \\ k_j & \text{otherwise} \end{cases}$$

where

$$k_j = \begin{cases} k^{sh} & \text{if } j = 1, \dots, 8 \\ k^{ax} & \text{if } j = 9, \dots, 16. \end{cases}$$

The label  $g$  is used to denote each damage scenario, ranging from 1 for the first floor up to 8 for the eighth floor; by convention,  $\mathbf{d}(0)$  refers to the undamaged case.

Before assessing the classification capability of the NN, a parametric analysis has been carried out to check the sensitivity to damage of the vibration frequencies. Table 1 collects the results regarding the horizontal motion; for the analysed system, the axial frequencies can be obtained by scaling the reported frequencies by a factor 10. Any considered damage state reduces all the frequencies, though the variation is rather limited even with a stiffness reduction by 25%, see Table 1. The capability to perform damage localization just by exploiting these data can be largely ineffective, since some trends in the table, such as the monotonic dependence of the frequencies of a vibration mode on the damage inter-story, can be hardly recognized.

As proposed in [2,25], the shape of the vibration modes—in particular that of the fundamental one in the case of a building featuring constant mass and stiffness at each story as for the case at hand—should be taken into account in the analysis, in order to localise and quantify damage. As previously remarked, employing FCN allows us not only to analyse separately each recorded signal, but also to exploit their interplay. Moreover, even if the sensitivity to damage of displacements in horizontal and vertical directions is the same, their joint use enabled by the FCN can lead to an improvement of the NN performances.

Due to the different range of values of vibration frequencies in the case of horizontal or vertical excitation of the structure, the axial response turns out to be richer in high-frequency vibrations. To correctly record the signals, the sampling rates have been set to 66.7 Hz to monitor the horizontal vibrations, and 667 Hz to monitor the vertical vibrations. Even for the higher vibration frequencies, output signals are assumed to be not distorted by the accelerometers: the transfer function of the sensor itself has to be very close to 1 for frequencies up to the mentioned values, so that the amplitude of sensor output very well matches the real structural response to be locally measured. If the structural vibration frequencies or the sampling rates get too close to the internal resonance frequency of the sensor, for some specific applications different, ad-hoc designed devices will be selected.

**Table 1 Shear vibration frequencies of the considered eight-story building, for the undamaged case (0) and under different damage scenarios, each one featuring a reduction by 25% of the stiffness at the inter-story corresponding to the scenario label**

Mode	Shear frequencies (Hz)								
	Damage scenario								
	0	1	2	3	4	5	6	7	8
1	1.175	1.131	1.134	1.140	1.146	1.154	1.162	1.169	1.173
2	3.484	3.368	3.427	3.480	3.467	3.400	3.355	3.375	3.445
3	5.678	5.522	5.668	5.570	5.467	5.614	5.647	5.488	5.522
4	7.673	7.521	7.630	7.401	7.663	7.448	7.532	7.600	7.394
5	9.409	9.287	9.143	9.322	9.126	9.362	9.114	9.396	9.105
6	10.825	10.745	10.471	10.764	10.665	10.490	10.812	10.528	10.598
7	11.873	11.833	11.659	11.510	11.750	11.857	11.596	11.560	11.755
8	12.516	12.505	12.456	12.378	12.273	12.227	12.328	12.421	12.484

In the analysis, each instance is made up by two multivariate time series, one for each excitation type, referring to different time intervals:  $I = [0, 10]s$  for the shear case and  $I = [0, 1]s$  for the axial case, respectively. Accordingly, the time series lengths are equal to  $L_0^1 = 667$  and to  $L_0^2 = 667$  for both the displacements in  $x$  and  $z$  direction. This benchmark has been exploited to test the FCN architecture with either one convolutional branch or two convolutional branches (see Figs. 2, 3). Indeed, what we are going to assess is the NN ability to perform the fusion of the information extracted through the concatenation layer, rather than the capacity to deal with time series of different lengths.

Two load types have been considered: first, we have excited the structure with lateral and vertical loads applied at each story and characterised by narrow frequency ranges, randomly sampled from an interval including, but not limited to, the structural frequencies; then we have applied, once again at each story, a white noise, assessing both the case in which all the shear frequencies have been excited, and the one in which just some of them have been covered by the noise frequency spectrum. With these two load types, we have been able to assess the NN performances in two different cases:

- case 1 (sinusoidal load case), in which the applied load is characterized by only few (a priori, random) frequencies;
- case 2 (white noise load case), in which the applied load is characterized by a higher number of (a priori, random) frequencies, lying in a given range.

This latter case corresponds to the one of random vibrations, for instance due to low-energy seismicity of natural or anthropic (urban) source [35], and is frequently adopted in literature, see e.g. [36]; the characteristic frequency range of seismic vibrations is site-dependent, being determined by the geographical and geological properties of the site. For example, in deep soft basins, the seismic vibrations are richer in low frequency components with respect to the ones in rock sites. For this reason, without any site characterization, it makes sense to assume more than a single frequency spectrum for the random vibrations.

#### **Case 1 (sinusoidal load case)**

In this first analysis, two different load combinations in the horizontal ( $x$ ) and vertical ( $z$ ) directions have been considered, to affect both the shear and axial vibration modes of the building. For each direction, the loads applied to the stories of the structure are given

**Table 2** Adopted random generation rules for the parameters  $\eta_l^{1sh}$  and  $\eta_l^{1ax}$  tuning the frequency and the magnitude of the applied sinusoidal load components in the x (a) and z (b) directions respectively

Parameter	Measurement unit	Adopted random generation rule
(a) x direction		
$f^{sh}$	Hz	{takerand ([1, 2.75, 4.5, 6.25, 8, 9.75, 11.5, 13.25, 15]) · (randn(0, $\sqrt{2}$ ))}
$\gamma^{sh}$	—	{ $\gamma_i^{dof} \cdot \text{randn}(0, 1)$ } with $\gamma_i^{dof} = \gamma^{dof}(i)$ and $\gamma^{dof} = [0.13, 0.25, 0.38, 0.50, 0.63, 0.75, 0.88, 1.00]$ and $i$ is the dof label
(b) z direction		
$f^{ax}$	Hz	{takerand ([10, 27.5, 45, 62.5, 80, 97.5, 115, 132.5, 150]) · (2randn(0, 1))}
$\gamma^{ax}$	—	{randn(0, 1)}

Here, we indicate with  $\text{randn}(0, \sigma)$  the sampling from a Gaussian probability distribution  $\mathcal{N}(0, \sigma^2)$ , where  $\sigma^2$  is its variance, and with  $\text{takerand}(\{v\})$  the uniform sampling from the discrete set of values  $\{v\}$

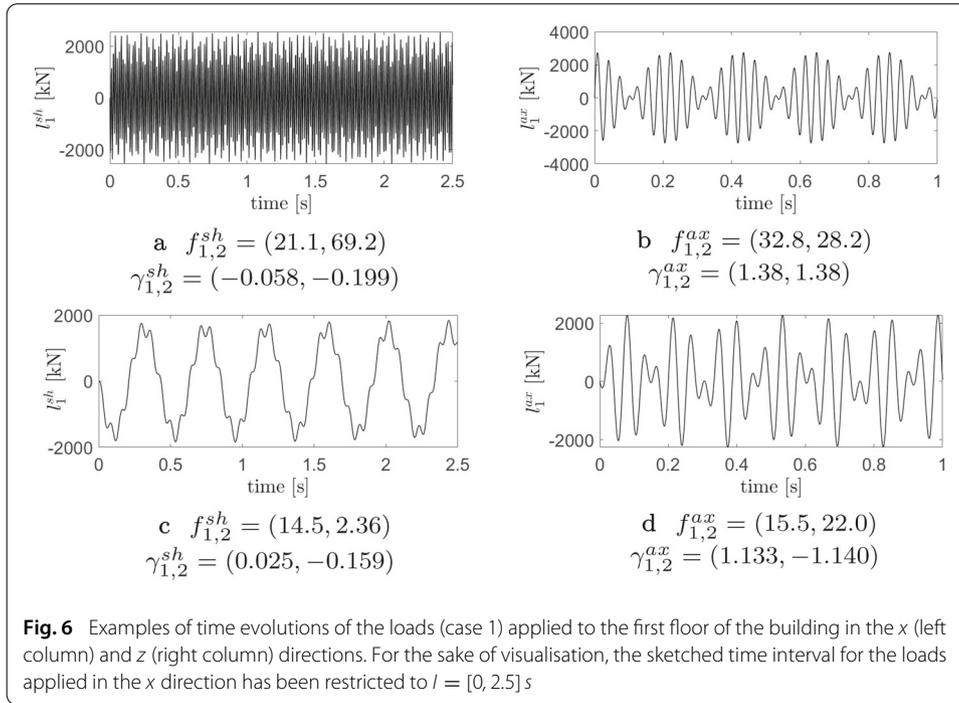
by the sum of two sinusoidal functions, whose amplitudes and time variations have been randomly generated. This expression for the load has been adopted to keep its description simple and, in comparison with single sinusoidal component case, to increase the set of frequencies that excite the structure. The applied load  $\mathbf{l} = [\mathbf{l}^{sh}, \mathbf{l}^{ax}]$  reads:

$$l_i^{sh}(t, \eta_l^{sh}) = \sum_{j=1}^2 F_i^{sh} \gamma_{i,j}^{sh} \sin(2\pi f_j^{sh} t), \quad i = 1, \dots, 8,$$

$$l_i^{ax}(t, \eta_l^{ax}) = \sum_{j=1}^2 F_i^{ax} \gamma_j^{ax} \sin(2\pi f_j^{ax} t), \quad i = 1, \dots, 8,$$

where:  $l_i^{sh}(t, \eta_l^{sh})$  and  $l_i^{ax}(t, \eta_l^{ax})$  are the amplitudes of the horizontal and vertical loads acting on the  $i$ -th floor;  $F_i^{sh} = 10^4$  kN and  $F_i^{ax} = 10^3$  kN are scaling parameters used to set the magnitude of the applied loads;  $\eta_l^{sh} = [\gamma^{sh}, f^{sh}]$  and  $\eta_l^{ax} = [\gamma^{ax}, f^{ax}]$ ;  $\gamma^{sh} \in \mathbb{R}$  and  $\gamma^{ax} \in \mathbb{R}$  are random scaling factors;  $f^{sh}, f^{ax} > 0$  set the frequencies of the sinusoidal components (see Table 2 for the adopted random generation rules).

The two sets of values adopted for the generation rule of  $f^{sh}$  and  $f^{ax}$  are chosen on the basis of the structural frequencies that could be excited both in the horizontal and vertical directions. At the same time, thanks to the adopted sampling rule,  $f^{sh}$  and  $f^{ax}$  may exceed these frequency ranges, producing instances in which the shear frequencies and/or the axial frequencies of the structure are not excited. Regarding the generation rule of the scaling parameter  $\gamma^{sh}$ , its dependency on the dofs of the structure through the factor  $\gamma^{dof}$  has been introduced in Table 2 in order to mimic the load distribution usually considered in a preliminary design process, when the shear behaviour of a regular building is evaluated. Keeping in mind that our principal interest here is to assess the prediction capacities of the NN architecture, this choice has enabled us to obtain displacement time series similar to the ones expected during the monitoring of the structure, although adopting a very simple generation rule for the applied lateral loads. Some examples of the time evolutions of the generated loads, applied to the first floor of the structure (hence of  $l_1^{sh}$  and  $l_1^{ax}$ ), are shown in Fig. 6.

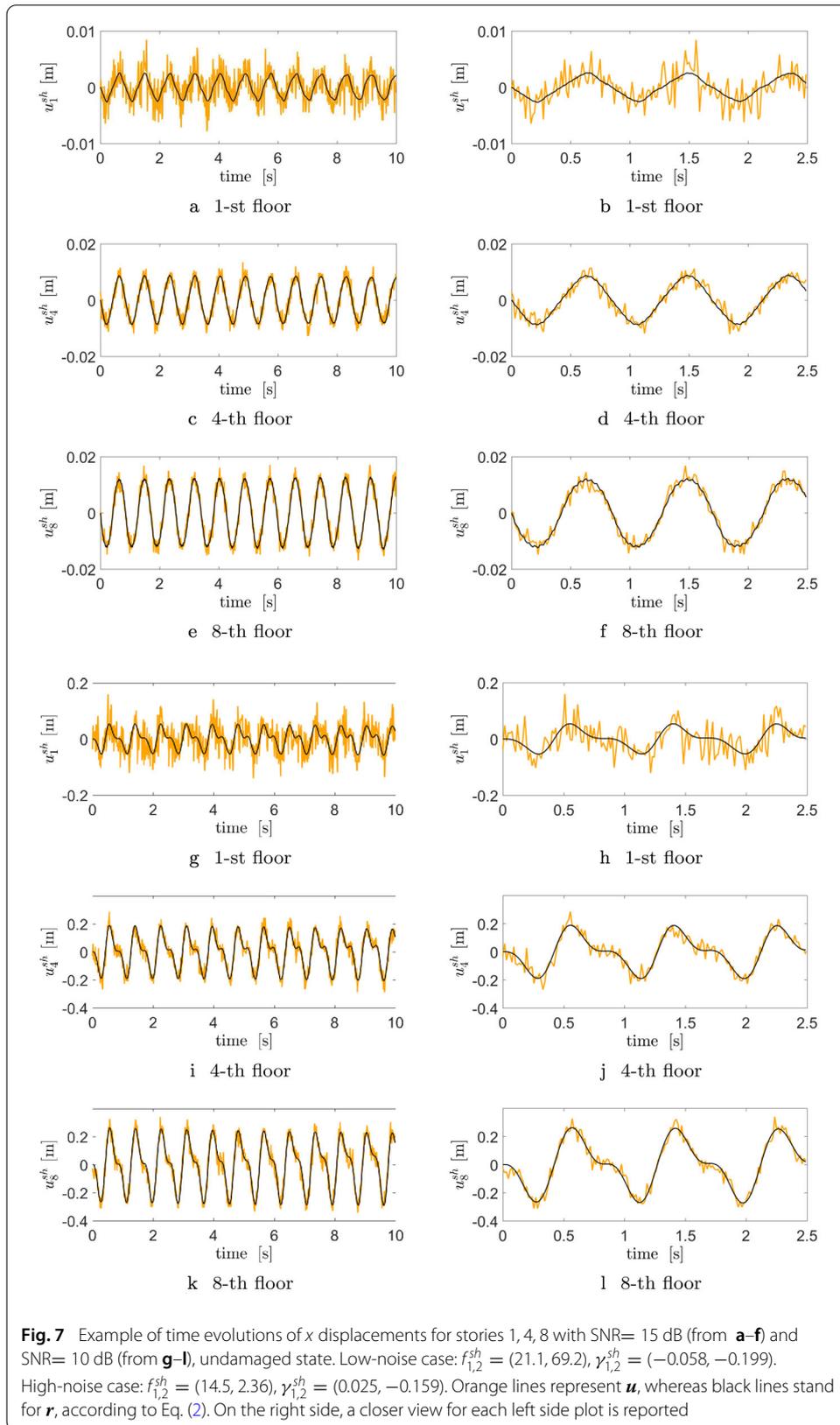


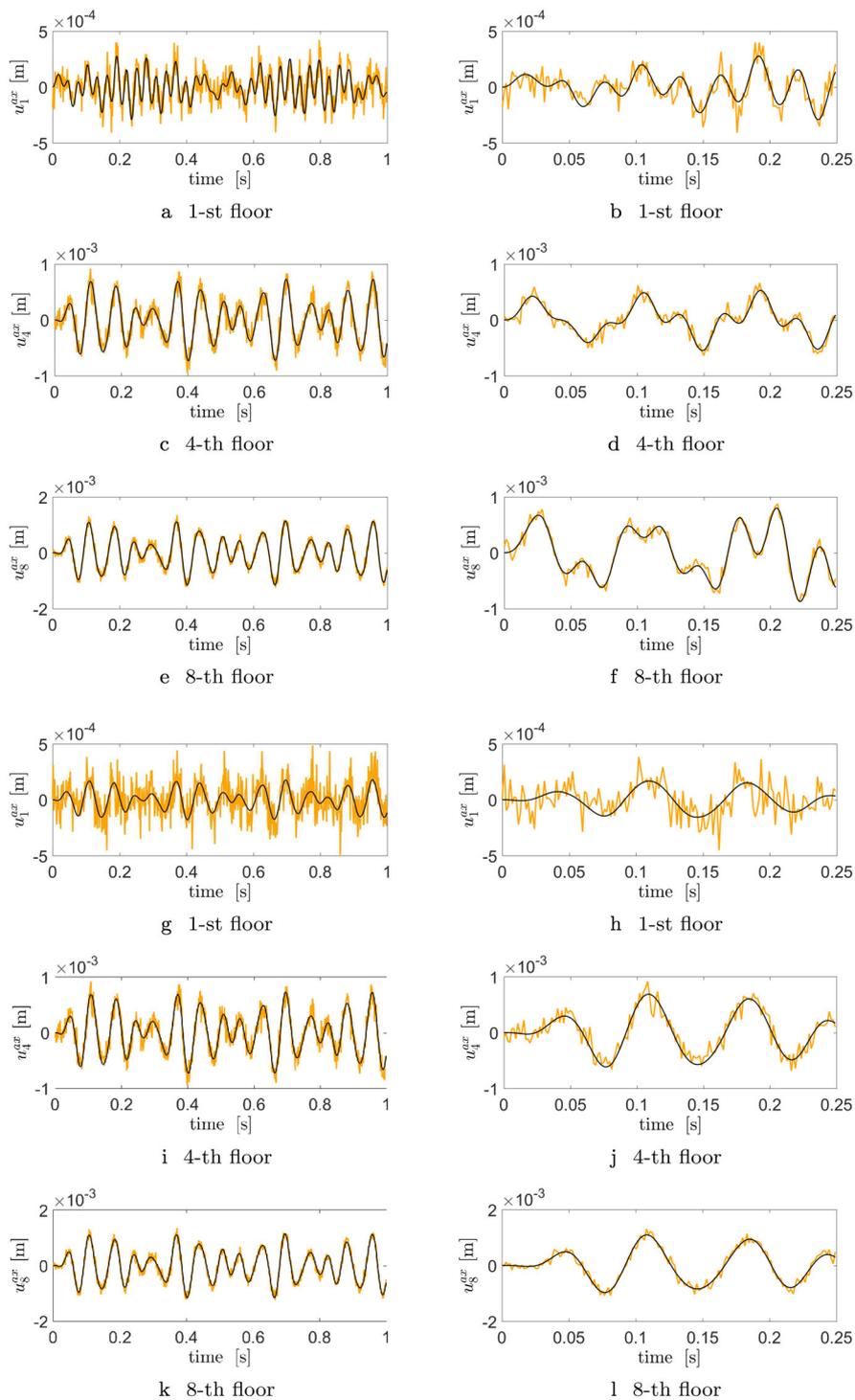
Through Eq. (2), we have added a measurement noise to mimic the output of a real monitoring system. For the sake of simplicity, the covariance matrix  $\Sigma_\epsilon \in \mathbb{R}^{16 \times 16}$  of such noise has been assumed to be diagonal, i.e.  $\Sigma_\epsilon = \sigma^2 \mathbb{I}$  where  $\sigma^2$  is the variance of the measurement error  $\epsilon$  in the horizontal and vertical directions for each floor, and  $\mathbb{I} \in \mathbb{R}^{16 \times 16}$  is the identity matrix.

Two sources of randomness have been assumed for the noise, due to environmental effects and to the transmission of the electrical signal. Their effects are superimposed in the covariance matrix with diagonal entries respectively amounting to  $\sigma_{env}^2$  and  $\sigma_{el}^2$ .

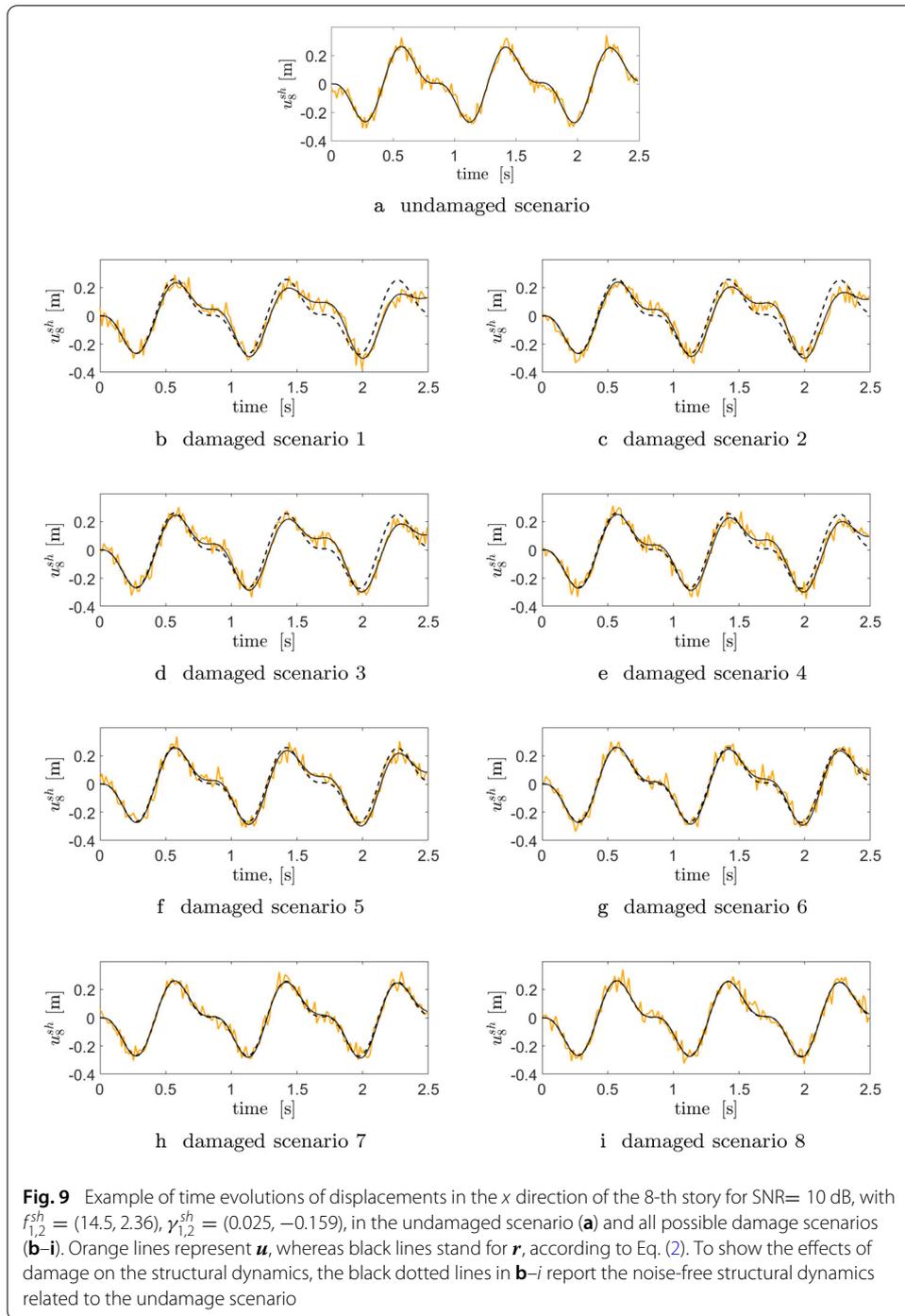
The environmental noise has been assumed to induce vibrations of the same amplitude and/or to affect in the same way the converted electrical signals, independently of the building floor. Given that horizontal motions at the top of the buildings are in general greater than displacements at the lower levels, this assumption leads small amplitude signals to be more affected, in relative terms, by the environmental noise. This is reasonable if we assume that the localised disturbances that arise because of the surrounding environment have the same magnitude independently of the building levels.

Regarding the electrical disturbance, the same noise level has been assumed both in directions  $x$  and  $z$ , in spite of the usually different technical specifications for sensors measuring displacements with different magnitude. This means that the electrical disturbances have the same effect, in statistical terms, on the measurement outcomes in the horizontal direction  $u_i^{sh}$  and in the vertical direction  $u_i^{ax}$ . Figures 7 and 8 respectively show examples of time evolutions of horizontal and vertical displacements, to highlight the effects of the above assumptions on the structural signals. These displacement components always refer to the undamaged case, and to the load conditions specified in the captions. According to what highlighted, it is noted that the displacements of the 8th story are less affected by noise than the ones of the 1st story.



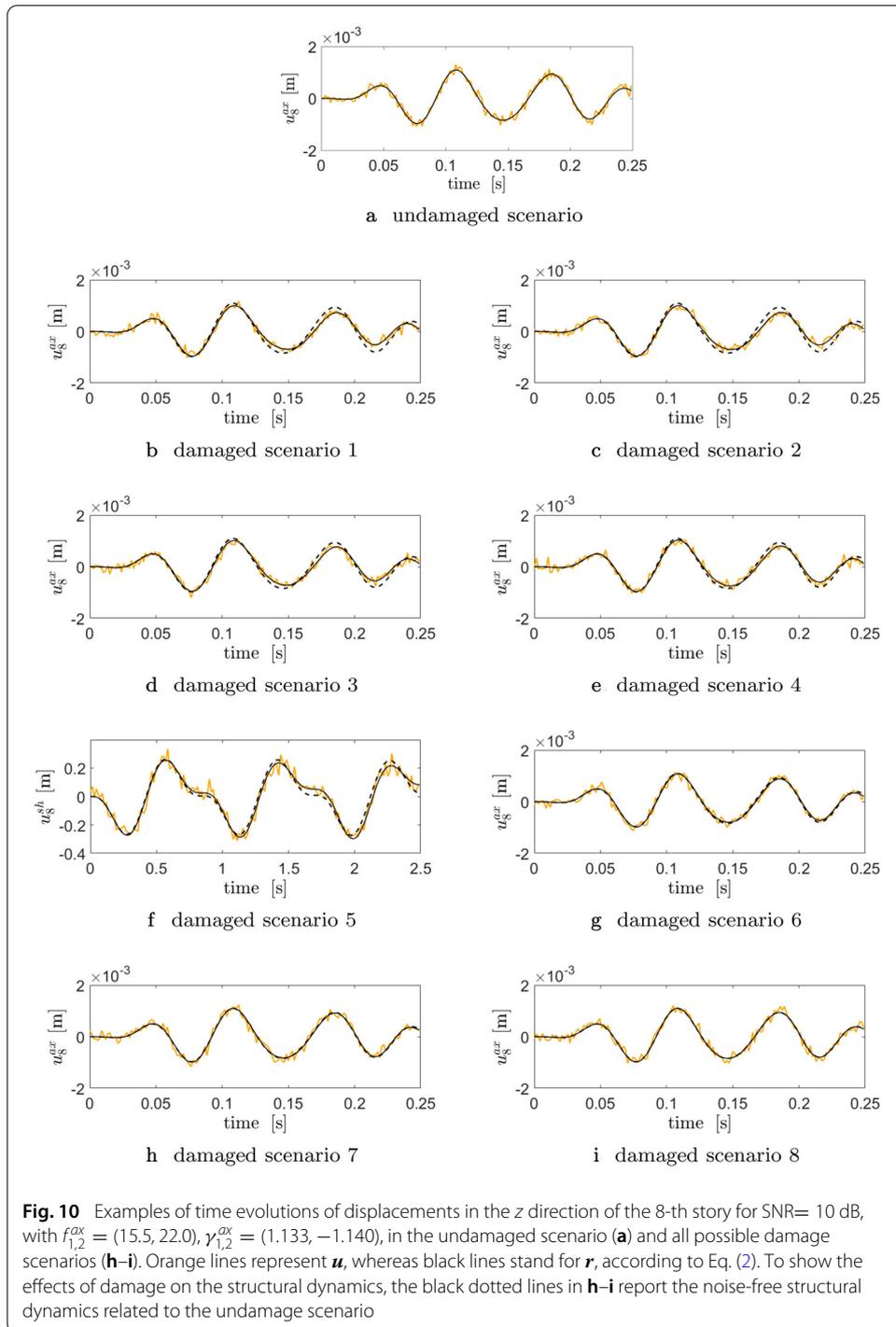


**Fig. 8** Example of time evolutions of  $z$  displacements for stories 1, 4, 8 with SNR= 15 dB (from **a-f**) and SNR= 10 dB (from **g-l**), undamaged state. Low-noise case:  $f_{1,2}^{ax} = (32.8, 28.2)$ ,  $\gamma_{1,2}^{ax} = (1.38, 1.38)$ . High-noise case:  $f_{1,2}^{ax} = (15.5, 22.0)$ ,  $\gamma_{1,2}^{ax} = (1.133, -1.140)$ . Orange lines represent  $\mathbf{u}$ , whereas black lines stand for  $\mathbf{r}$ , according to Eq. (2). On the right side, a closer view for each left side plot is reported

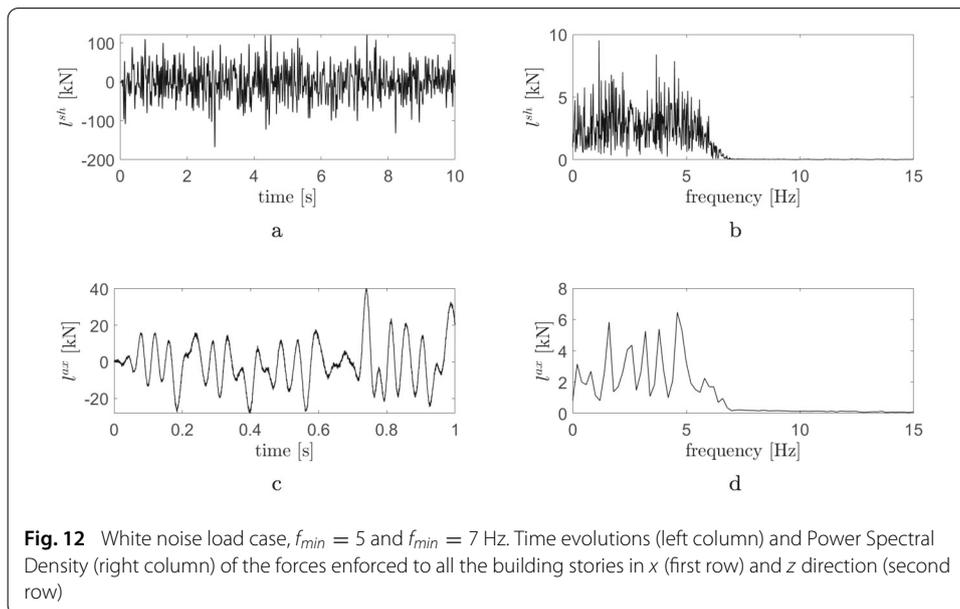
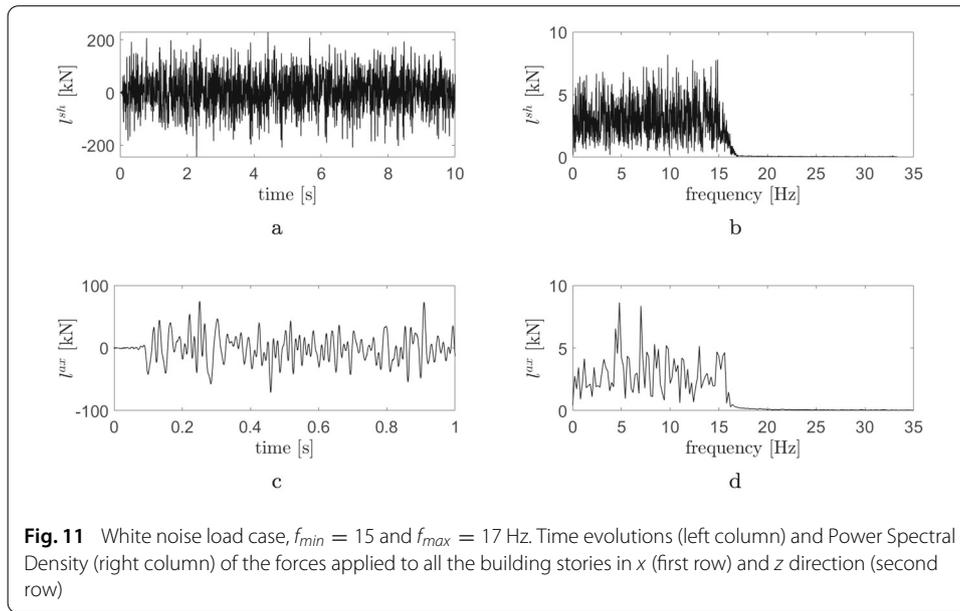


Due to the random generation of the applied load, different structural frequencies are excited in each simulation. To provide different scenarios also in terms of sensor accuracy (see also [37]), two levels of SNR of 15 dB and 10 dB have been adopted. The SNR is a summary indicator, referring to the overall level of noise corruption for the displacements in one direction. Still referring to Figs. 7 and 8, differences in terms of corruption levels between the two sensor accuracy scenarios can be appreciated.

To build the dataset required for the NN training, the procedure described so far has been adopted for all the damage scenarios. Figures 9 and 10 respectively show the effects



of damage on  $u_8^{sh}$  and  $u_8^{ax}$ , highlighting the sensitivity of this output to the handled damage state. To better highlight this sensitivity, the time evolutions in Figs. 9 and 10 are provided for  $I = [0, 2.5]$ s and  $I = [0, 0.25]$ s only, even though  $I = [0, 10]$ s and  $I = [0, 1]$ s have been adopted for the NN training. Drifts from the responses relevant to the undamaged case can be observed when the damage scenarios refer to the stiffness reduction of the lowest stories; however, it looks nearly impossible, in general, to perform any classification of the damage scenarios without any effectively trained classifier.



**Case 2 (white noise load case)**

In the second load case we have accounted for random vibrations caused e.g. by low-energy seismicity [36]. The applied loads  $\mathbf{l} = [\mathbf{l}^{sh}, \mathbf{l}^{ax}]$ , with  $i = 1, \dots, 8$ , at each floor and each time instants are obtained by first sampling out the values from a normal distribution  $\mathcal{N}(0, 10^4)$  and then low-pass filtering them with a “roll-off” set between frequencies  $f_{min}$  and  $f_{max}$ . Two different scenarios have been considered for the frequency range of the applied excitations:  $f_{min} = 15$  and  $f_{max} = 17$  Hz;  $f_{min} = 5$  and  $f_{max} = 7$  Hz. In the first case all the shear modes and the first axial mode have been excited; in the second case, just the first three shear modes and no axial frequencies have been excited, see Table 1. Figures 11 and 12 respectively provide an overview of the simulated forces for the two cases.

### Dataset composition and NN training

We now detail the construction of the employed datasets and the NN training phase. Each of the two classifiers has been trained on a different dataset, made by instances generated by evaluating the physics-based model for different loading and damage conditions. Each instance is made up by  $N_0 = 16$  time series recordings of displacements (in the two directions, for each of the 8 floors) of length  $L_0 = 667$ . Due to the assumed shear-type behavior of the building, all the points belonging to each rigid floor share the same accelerations and displacements; hence, there is no need to plug-in specific optimal strategies to locate sensors in the network, which could be instead of interest in case of very localized damage events breaking the validity of the rigid floor assumption.

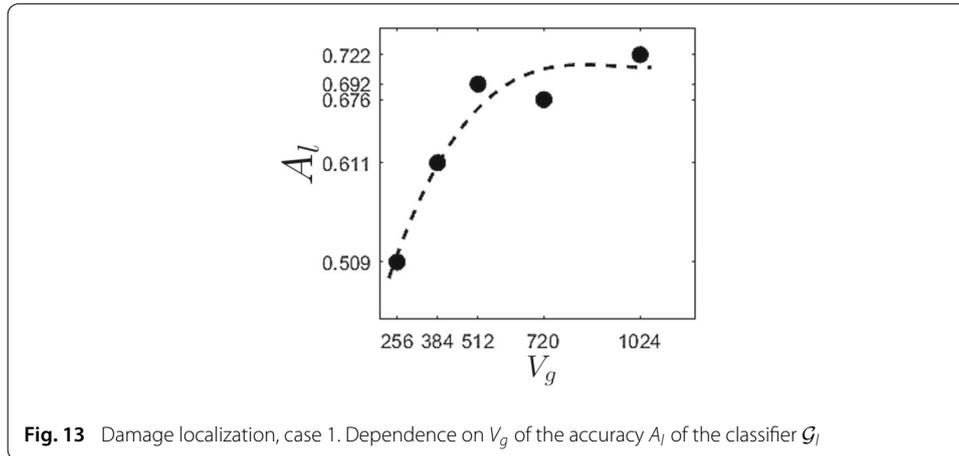
Two global datasets  $\mathbb{D}^d$  and  $\mathbb{D}^l$  made by  $V = 4608$  instances each have been generated, and then split onto a training, a validation and a testing set, thus yielding  $\mathbb{D}^d = \mathbb{D}_{train}^d \cup \mathbb{D}_{val}^d \cup \mathbb{D}_{test}^d$  and  $\mathbb{D}^l = \mathbb{D}_{train}^l \cup \mathbb{D}_{val}^l \cup \mathbb{D}_{test}^l$ , with  $V = V^{train} + V^{val} + V^{test}$  in both cases.

For the splitting of the dataset  $\mathbb{D}^d$  into training  $\mathbb{D}_{train}^d$ , validation  $\mathbb{D}_{val}^d$  and test  $\mathbb{D}_{test}^d$  sets, no specific rules are available, and only some heuristics can be used – see, e.g., [27]. We have thus employed 75% of  $V$  to train and validate the NN ( $V^{train}$  and  $V^{val}$ ), and the remaining 25% ( $V^{test}$ ) to test it. Within the first subset, 75% of the instances have been in turn allocated for training, and the remaining 25% for validation. The final dataset subdivision then reads:  $V^{train} = 56.25\%V$ ,  $V^{val} = 18.75\%V$ , and  $V^{test} = 25\%V$ . The splitting of  $\mathbb{D}^l$  has been done identically. The large number of instances employed for validation and test has allowed us to perform a robust assessment of the NN generalization capabilities. This has been done without limiting the information content that can be employed for the NN training; in fact, the dataset dimensions can be arbitrarily enlarged, if necessary, through a synthetic generation of the new instances, still keeping the same proportions.

During the training, an equal number of instances  $V_g^{train} = V^{train}/G$  related to each damage scenario  $g = 0, \dots, 8$  (the undamaged case has been considered, too, in addition to the  $G = 8$  possible cases of damage) have been provided to the NN, to avoid the construction of a biased dataset  $\mathbb{D}_{train}^d$ ; the same has been done for  $\mathbb{D}_{train}^l$ . In this way, we indeed prevent the NN to be prone to return the class labels that have been more frequently presented in the training stage.

There are no specific rules to set  $V_g^{train}$  (and, therefore, the overall dimension  $V_g = V/G$  of simulated cases for each damage scenario) a priori. Only few theoretical studies provided some recommendations for specific cases, see, e.g., [38]; however, they are not applicable to FCNs. In general, the problem complexity and the employed NN architecture must be taken into account on a case-by-case basis. For this reason, we have evaluated the  $\mathcal{G}_d$  and  $\mathcal{G}_l$  classifiers accuracies  $A_d$  and  $A_l$  on the validation set  $\mathbb{D}_{val}^d$  and  $\mathbb{D}_{val}^l$ , and the training time at varying  $V_g^{train}$ . We have then chosen the best dataset size according to a tradeoff between the two aforementioned indicators, and keeping in mind that the time required to generate a dataset and to train the NN both scale linearly with  $V_g^{train}$ . The  $\mathcal{G}_d$  classifier accuracy is defined as the ratio  $A_d = V_{\star}^{val}/V^{val}$ , where  $V_{\star}^{val}$  is the number of instances of  $\mathbb{D}_{val}^d$  which are correctly classified by  $\mathcal{G}_d$ ; the  $\mathcal{G}_l$  classifier accuracy  $A_l$  is defined in a similar way.

Let us now see how we have determined the overall dataset size  $V$  by applying the heuristic approach previously discussed. In Fig. 13, the accuracy  $A_l$  at varying values of  $V_g$  is reported, by considering the local case 1. By increasing  $V_g$  from 256 to 384,  $A_l$  is



highly affected, while a further increasing yields a smaller gain in accuracy. The non-monothonic variation of  $A_l$  with respect to  $V_g$  is due to the randomness of the procedure, and in particular to the initialization of the weights of the convolutional filters. For the above reasons, we have adopted  $V_g = 512$  during the training phase.

Treating the damage detection task for case 1, a total number of  $V = 9216$  instances have been generated. Half of the instances refers to the undamaged conditions, half to damaged conditions. Each damage scenario is equally represented ( $V_g = 512$  instances each). Regarding instead the damage localization task,  $V = 4608$  and  $V_g = 512$  (including the undamaged case  $g = 0$ ) have been adopted.

Still adopting the discussed heuristic criterion for the determination of the overall dataset dimension,  $V = 4096$  has been used for the damage detection task when the white noise load case is treated. Once again, half of the instances refers to the undamaged conditions, half to the damage condition. Each damage scenario is equally represented ( $V_g = 128$  instances each). Regarding the damage localization task,  $V = 4608$  and  $V_g = 128$  (including the undamaged case  $g = 0$ ) have been adopted.

### Classification outcomes

We now report the numerical results obtained for the two load cases, and for the two required tasks of damage detection and damage localization. The obtained classification outcomes are affected by the NN architecture, either with one or two convolutional branches, depending on whether the horizontal and vertical sensing are both considered or not. In particular, when treating the damage localization task in presence of the white noise load condition, we will also try to assess the impact of each input channel  $\mathcal{F}_0^n$  on the overall NN accuracy.

Useful indications about the quality of the training can be derived from the behavior of the loss functions  $J_d(\mathbf{Y}, \mathbf{p})$  and  $J_l(\mathbf{Y}, \mathbf{p})$ —see Eq. (6)—of  $\mathcal{G}_d$  and  $\mathcal{G}_l$ , and of the accuracies  $A_d$  and  $A_l$  on the training and validation sets ( $\mathbb{D}_{train}^t$  and  $\mathbb{D}_{val}^t$  for  $\mathcal{G}_d$ ;  $\mathbb{D}_{train}$  and  $\mathbb{D}_{val}$  for  $\mathcal{G}_l$ ) as a function of the number of iterations. This latter depends on both the number of epochs and the minibatch size chosen for the training.<sup>3</sup>

<sup>3</sup> In other words, if the dataset is composed by 100 instances and a minibatch size of 10 instances is adopted, after the first epoch the iteration number is equal to 10.

**Table 3 Damage detection, case 1**

SNR (dB)	$\{\mathcal{F}_*\}$	$A_d$
15	$\{\mathbf{u}_T^{sh}\}_{i=1}^8$	0.814
15	$\{\mathbf{u}_T^{ax}\}_{i=1}^8$	0.850
15	$\{\mathbf{u}_T^{sh}\}_{i=1}^8$ and $\{\mathbf{u}_T^{ax}\}_{i=1}^8$	0.879
10	$\{\mathbf{u}_T^{sh}\}_{i=1}^8$	0.768
10	$\{\mathbf{u}_T^{ax}\}_{i=1}^8$	0.775
10	$\{\mathbf{u}_T^{sh}\}_{i=1}^8$ and $\{\mathbf{u}_T^{ax}\}_{i=1}^8$	0.765

Accuracy  $A_d$  of the classifier  $\mathcal{G}_d$  evaluated on  $\mathbb{D}_{test}^d$

To evaluate the NN performances, the adopted indices are still  $A_d$  and  $A_l$ , yet evaluated on  $\mathbb{D}_{test}^d$  and  $\mathbb{D}_{test}^l$ . These indices are always compared against the ones produced by a random guess, equal to 0.5 for  $\mathcal{G}_d$ , and to  $1/9 = 0.111$  for  $\mathcal{G}_l$ . For the damage localization case, the misclassification is measured by a confusion matrix in which the rows correspond to the target classes and the columns to the NN predictions.

#### **Damage detection and localization in case 1—sinusoidal load case**

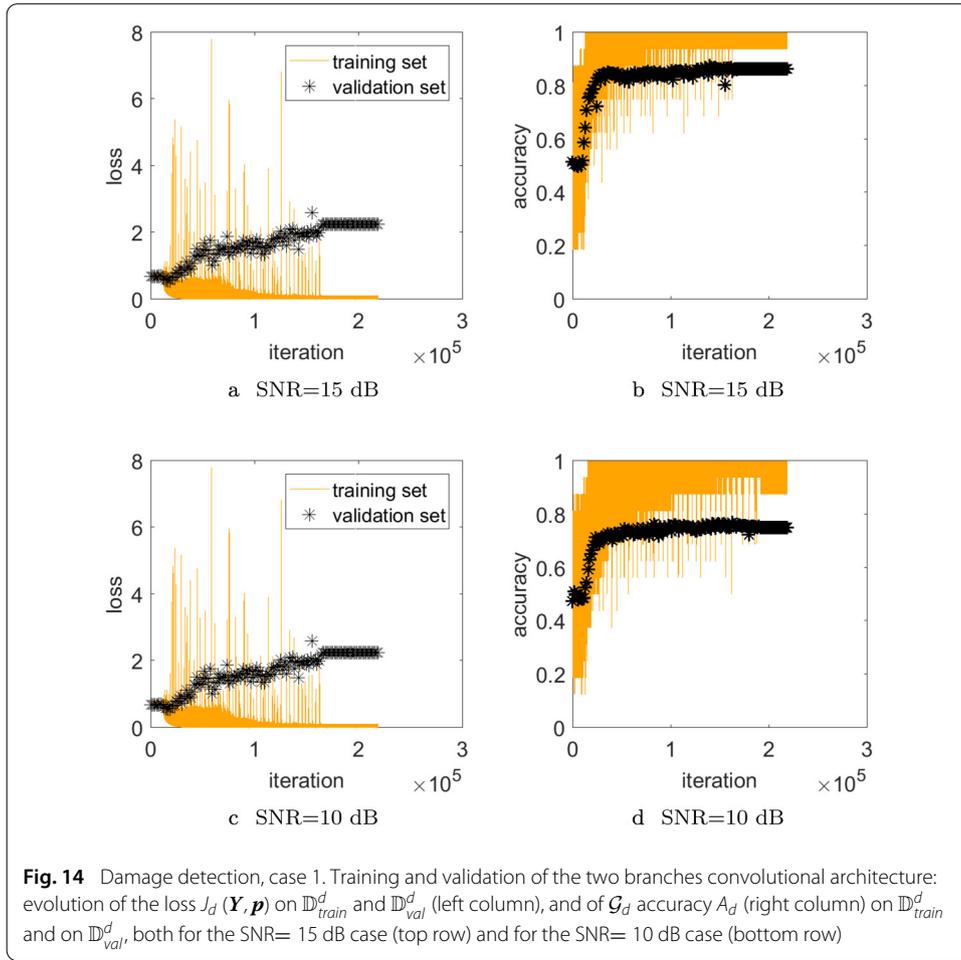
In Table 3 the accuracies  $A_d$  of  $\mathcal{G}_d$  on  $\mathbb{D}_{test}^d$  for the two considered noise levels (SNR= 15 dB and SNR= 10 dB) are reported. NN architectures with both one and two convolutional branches have been tested.

The classifier  $\mathcal{G}_d$  reaches  $A_d = 0.879$  for SNR= 15 dB and  $A_d = 0.775$  for SNR= 10 dB. These outcomes obtained on high-noise datasets show the potentialities of the proposed approach in view of facing real engineering applications. Indeed, noise effect is a principal concern especially when pervasive and low-cost microelectromechanical systems (MEMS) sensor networks are employed [39], so that the possibility to handle it through FCNs may enhance the application of MEMS networks. Moreover, thanks to our procedure, we have been able to avoid the data pre-processing required by any ML approach based on problem specific features.

Figure 14 reports the evolution of the training and validation loss for the dataset with SNR= 15 dB and SNR= 10 dB. The iteration number accounts for the number of times the NN weights are modified during the training process. The depicted training and validation loss functions refer to the case in which a two branches convolutional architecture has been employed to detect damage. The several spikes observed both in the loss and accuracy graphs are due to the stochastic nature of the training algorithm. During the early stages of the training, the NN displays the most significant gains in terms of classification accuracy, while further increasing the number of iterations only yields a limited effect on the generalization capabilities of the NN. Due to the lack of improvements, the early-stopping criterion has finally terminated the training.

Moving to the damage localization task, Table 4 collects the results related to the outcomes of  $\mathcal{G}_l$  on  $\mathbb{D}_{test}^l$  obtained for two different noise levels.

The results show that the NN performances benefit from the employment of a two branches architectures:  $A_l$  increases, compared to the best outcome of the single convolutional layer architecture, from 0.769 to 0.812 for the SNR= 15 dB case, and from 0.654 to 0.707 for the SNR= 10 dB case. This means that the NN has succeeded in performing a data fusion of the extracted information for the sake of classification.



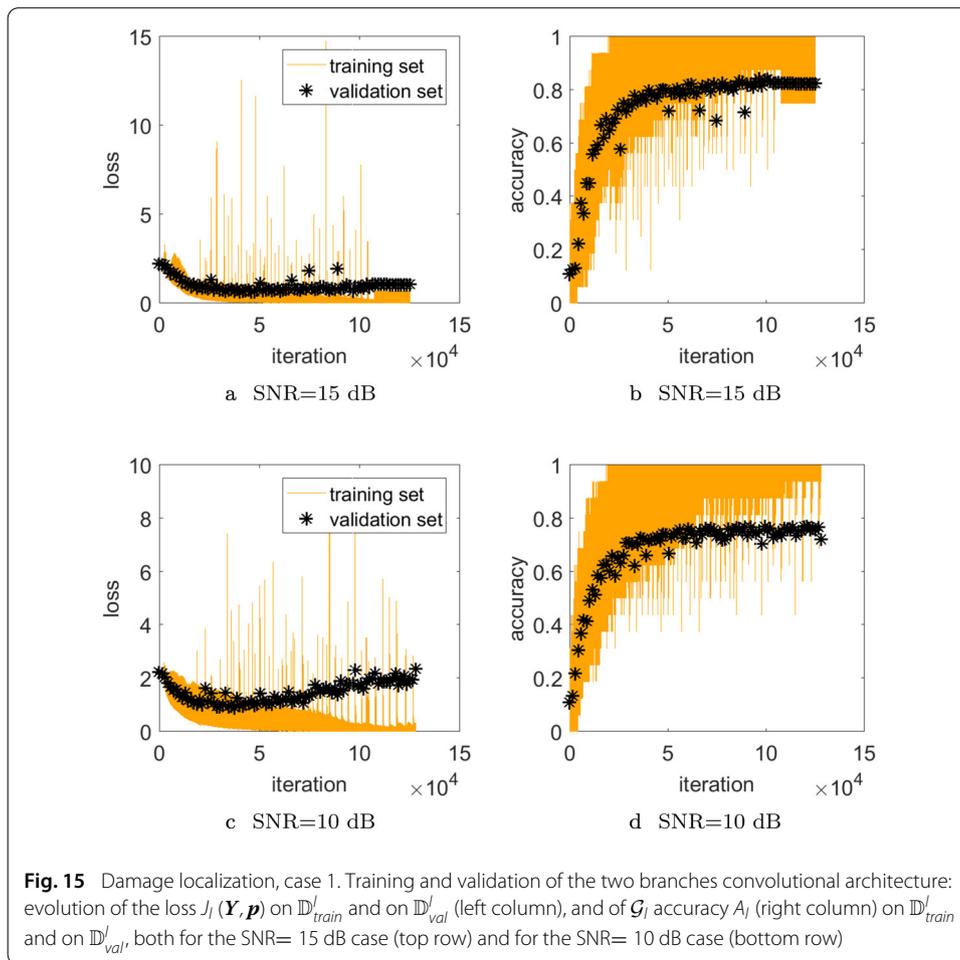
**Table 4** Damage localization, case 1

SNR (dB)	$\{\mathcal{F}_*\}$	$A_l$
15	$\{\mathbf{u}_i^{sh}\}_{i=1}^8$	0.768
15	$\{\mathbf{u}_i^{ax}\}_{i=1}^8$	0.769
15	$\{\mathbf{u}_i^{sh}\}_{i=1}^8$ and $\{\mathbf{u}_i^{ax}\}_{i=1}^8$	0.812
10	$\{\mathbf{u}_i^{sh}\}_{i=1}^8$	0.654
10	$\{\mathbf{u}_i^{ax}\}_{i=1}^8$	0.642
10	$\{\mathbf{u}_i^{sh}\}_{i=1}^8$ and $\{\mathbf{u}_i^{ax}\}_{i=1}^8$	0.707

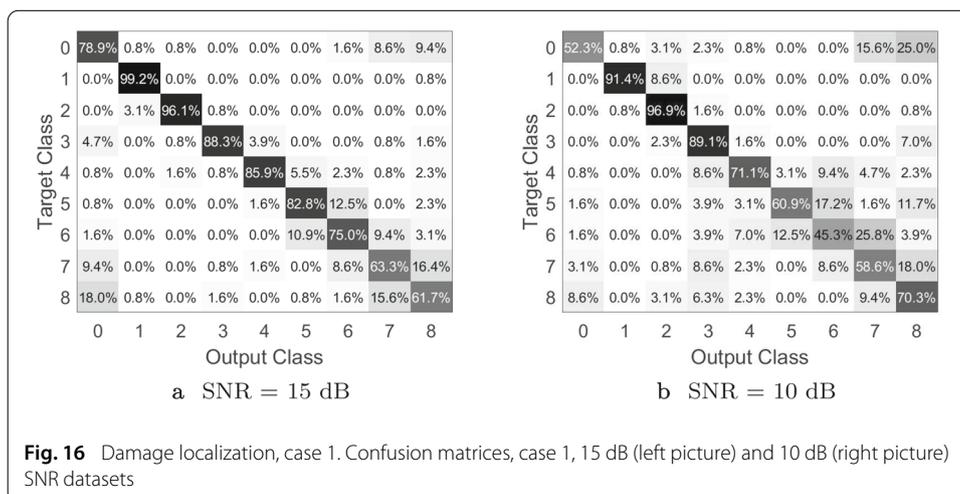
Accuracy  $A_l$  of the classifier  $\mathcal{G}_l$  evaluated on  $\mathbb{D}_{test}^l$

The values of  $A_d$  and  $A_l$  are quite close, despite the greater complexity of the damage localization problem; this might be due to the intrinsic capability of the FCN to detect correlations between different sensor recordings, allowing us to perform a correct damage localization.

Figure 15 reports the evolution of the training and validation loss functions on  $\mathbb{D}_{val}^l$  and  $\mathbb{D}_{test}^l$  for the datasets with SNR= 15 dB and SNR= 10 dB, in the case where a two branches convolutional architecture has been employed. Compared with Fig. 14, a smaller difference in terms of loss and accuracy can be highlighted. This is due to the greater complexity of the damage localization task, that requires to exploit the computational resources of the NN entirely. Indeed, the same number of filters  $N_1$ ,  $N_2$  and  $N_3$  has been used for both the



**Fig. 15** Damage localization, case 1. Training and validation of the two branches convolutional architecture: evolution of the loss  $J_l(\mathbf{Y}, \mathbf{p})$  on  $\mathbb{D}_{train}^l$  and on  $\mathbb{D}_{val}^l$  (left column), and of  $\mathcal{G}_l$  accuracy  $A_l$  (right column) on  $\mathbb{D}_{train}^l$  and on  $\mathbb{D}_{val}^l$ , both for the SNR= 15 dB case (top row) and for the SNR= 10 dB case (bottom row)



**Fig. 16** Damage localization, case 1. Confusion matrices, case 1, 15 dB (left picture) and 10 dB (right picture) SNR datasets

classification tasks, in spite of their different complexity. On the other hand, we expect that  $A_d$  on  $\mathbb{D}_{test}^d$ , reported in Table 3, would not be affected by reducing the number of filters. This conclusion can be reached by looking at Fig. 14 and observing that, during the last stages of the training,  $A_d$  on  $\mathbb{D}_{train}^d$  is shown to be always greater than the one obtained on  $\mathbb{D}_{val}^d$ .

**Table 5 Damage detection, case 2**

$f_{min} - f_{max}$ (Hz)	$\{\mathcal{F}_*\}$	$A^d$
15 – 17	$\{u_i^{sh}\}_{i=1}^8$	0.998
15 – 17	$\{u_i^{ax}\}_{i=1}^8$	0.997
15 – 17	$\{u_i^{sh}\}_{i=1}^8$ and $\{u_i^{ax}\}_{i=1}^8$	0.999
5 – 7	$\{u_i^{sh}\}_{i=1}^8$	0.996
5 – 7	$\{u_i^{ax}\}_{i=1}^8$	0.892
5 – 7	$\{u_i^{sh}\}_{i=1}^8$ and $\{u_i^{ax}\}_{i=1}^8$	0.998

Accuracy  $A_d$  of the classifier  $\mathcal{G}_d$  evaluated on  $\mathbb{D}_{test}^d$

**Table 6 Damage localization, case 2**

$f_{min} - f_{max}$ (Hz)	$\{\mathcal{F}_*\}$	$A^l$
15 – 17	$\{u_i^{sh}\}_{i=1}^8$	0.972
15 – 17	$\{u_i^{ax}\}_{i=1}^8$	0.972
15 – 17	$\{u_i^{sh}\}_{i=1}^8$ and $\{u_i^{ax}\}_{i=1}^8$	0.986
5 – 7	$\{u_i^{sh}\}_{i=1}^8$	0.993
5 – 7	$\{u_i^{ax}\}_{i=1}^8$	0.892
5 – 7	$\{u_i^{sh}\}_{i=1}^8$ and $\{u_i^{ax}\}_{i=1}^8$	0.972

Accuracy  $A_l$  of the classifier  $\mathcal{G}_l$  evaluated on  $\mathbb{D}_{test}^l$

In Fig. 16, the confusion matrices related to the two datasets (SNR= 15 dB and SNR= 10 dB) are reported. Most of the errors concern the classification of the damage scenarios in which the inter-story stiffness of the highest floors has been reduced, as shown by the entries of the 7-th and 8-th rows and columns of the matrices. This outcome is not surprising if we consider that these damage scenarios only induce small variations in the shear vibration frequencies. Moreover, by looking at Figs. 9 and 10, we can remark that the time evolution of the structural motions under these damage scenarios cannot be easily distinguished from the undamaged case.

#### Damage detection and localization in case 2

We now consider the outcomes of the trained classifiers in the case where a random disturbance is applied to the structural system. Regarding the damage detection task, with this type of excitation the NN is able to distinguish between undamaged and damaged instances almost perfectly (see Table 5). Indeed,  $A_d = 0.999$  and  $A_d = 0.998$  have been reached by the two convolutional branches architecture when  $f_{min} = 15$  and  $f_{max} = 17$  Hz, or  $f_{min} = 5$  and  $f_{max} = 7$  Hz, have been selected as frequency ranges for the applied lateral and vertical forces.

We next consider the NN outcomes for the damage localization task. With this type of excitation, the NN is able to accomplish an extremely accurate classification of the damaged scenarios, reaching  $A_l = 0.986$  and  $A_l = 0.993$  when  $f_{min} = 15$  and  $f_{max} = 17$  Hz or  $f_{min} = 5$  and  $f_{max} = 7$  Hz have been used, respectively. In the former case, the best classification performances have been obtained by the two convolutional branches architecture, as shown in Table 6. For the latter case, the NN employing as input  $\mathcal{F}_* = \{u_i^{sh}\}_{i=1}^8$  provides the best classification result. The better performances of the NN employing  $\mathcal{F}_* = \{u_i^{sh}\}_{i=1}^8$  rather than  $\mathcal{F}_* = \{u_i^{ax}\}_{i=1}^8$  is likely due to the fact in this latter case no axial frequencies have been excited by the applied load, as remarked in Case 2 (white noise load case). However, this fact also shows that the data fusion operated by the two convolutional

**Table 7 Damage localization, case 2**

$N_0$	$\{\mathcal{F}_*\}$	$A_l$
1	$i = 8$	0.226
2	$i = (4, 8)$	0.722
3	$i = (2, 4, 8)$	0.774
4	$i = (2, 4, 6, 8)$	0.906
5	$i = (2, 4, 6, 7, 8)$	0.865
6	$i = (2, 4, 5, 6, 7, 8)$	0.937
7	$i = (2, 3, 4, 5, 6, 7, 8)$	0.899
8	$i = (1, 2, 3, 4, 5, 6, 7, 8)$	0.993

Accuracy  $A_l$  of the classifier  $\mathcal{G}_l$  evaluated on  $\mathbb{D}_{test}^l$ . Different numbers  $N_0$  of input channels  $\mathcal{F}_*$ , related to  $u_i^{sh}$ , are employed. Here,  $f_{min} = 5$  and  $f_{max} = 7$  Hz

branches architecture has been only partially able to select the most important information required for the damage localization task. Nevertheless, very good results have been reached by also employing  $\mathcal{F}_* = \{u_i^{ax}\}_{i=1}^8$  (see Table 6).

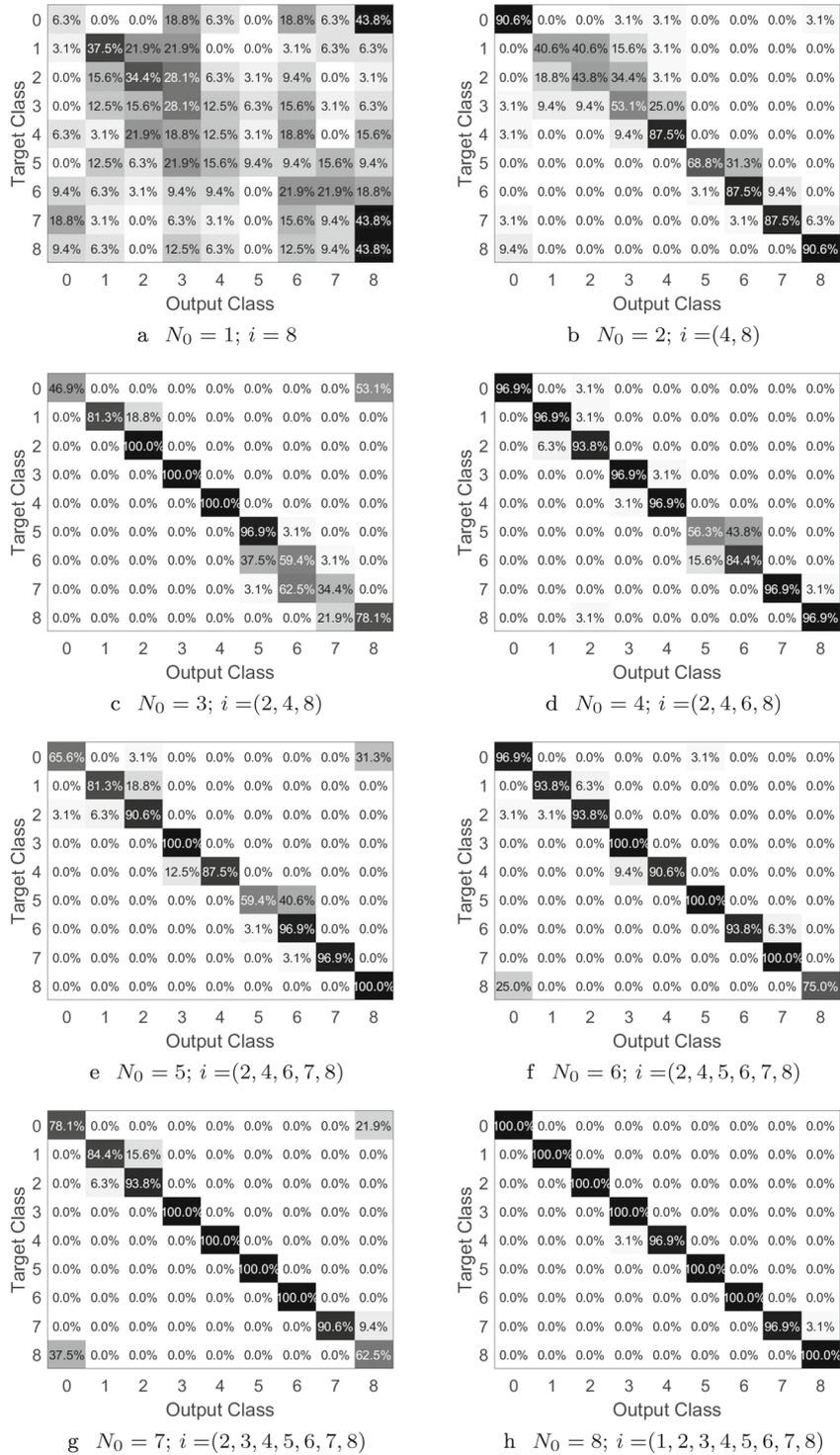
We highlight the effect of each incoming signal on the classification outcomes (see Table 7), since the accuracy  $A_l$  on  $\mathbb{D}_{test}^l$  changes for different numbers of input signals  $N_0$ . The results refer to the case in which only some of the displacements  $u_i^{sh}$ ,  $i = 1, \dots, 8$  have been considered, and  $f_{min} = 5$  and  $f_{max} = 7$  Hz. The corresponding confusion matrices are sketched in Fig. 17, showing that the classification error related to a damage scenario  $g$  is reduced when the corresponding  $u_g^{sh}$ , that is the signal acquired on the floor whose inter-story stiffness has been reduced, is used as input for the NN.

## Conclusions

In this paper, we have investigated a new strategy for real-time structural health monitoring, treating damage detection and localization as classification tasks [3], and framing the proposed procedure in the family of SBC approaches [4]. We have proposed to employ fully convolutional networks to analyse time series coming from a set of sensors. Fully convolutional networks architectures differing for the number of convolutional branches have been exploited to deal with datasets including time signals of different length and sampling rate. Convolutional layers have been shown to enable the automatic extraction of features to be used for the classification task at hand. The neural network architecture has been trained in a supervised manner on data generated through the numerical solution of a physics-based model of the monitored structure under different damage scenarios.

In the considered numerical benchmarks, we have obtained extremely good performances concerning both damage detection and damage localization, even in the presence of noise, when the applied loads can be characterized either (i) in terms of a few (a priori, random) frequencies, or (ii) by a higher number of frequencies, within a given range. Especially in the second case, the outcomes of the NN classifier have shown the potentialities of the proposed procedure in view of the application to real-life cases.

In future works, we aim to employ the proposed architecture to deal with data coming from real monitoring systems, tackling the main limit of the proposed procedure concerning the mimicking of the real structural response. This is a well-known problem in the machine learning community [40]. By coupling recurrent layers branches to the proposed convolutional ones, we expect to further increase the NN performances. As further steps, we will try to exploit model order reduction techniques for the dataset construc-



**Fig. 17**  $G_I$  confusion matrices for different number  $N_0$  of input channels  $\mathcal{F}_*^i$ . Case 2,  $f_{min} = 5$  and  $f_{max} = 7$  Hz

tion, extending the proposed methodology to more complex structural configurations and damage scenarios, and to design the set of sensors according to a Bayesian optimization technique [20,41,42].

#### Acknowledgements

The authors thank Andrea Opreni (Politecnico di Milano) for fruitful discussions about DL architectures. LR, SM and AC gratefully acknowledge the financial support from MIUR Project PRIN 15-2015LYYA 8 "Multi-scale mechanical models for the design and optimization of microstructured smart materials and metamaterials".

#### Authors' contributions

The authors contributed equally to this work. All authors read and approved the final manuscript.

#### Availability of data and materials

Both the numerical benchmark and the neural network architecture have been exhaustively described. The reader can verify the performance of the proposed method by running analogous numerical experiments.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy, <sup>2</sup>MOX, Dipartimento di Matematica, Politecnico di Milano, Piazza L. da Vinci 32, 20133 Milano, Italy.

Received: 8 November 2019 Accepted: 25 August 2020

Published online: 24 September 2020

#### References

- Chang PC, Flatau A, Liu SC. Health monitoring of civil infrastructure. *Struct Health Monit.* 2003;2(3):257–67. <https://doi.org/10.1177/1475921703036169>.
- Eftekhar Azam S, Mariani S. Online damage detection in structural systems via dynamic inverse analysis: a recursive bayesian approach. *Eng Struct.* 2018;159:28–45. <https://doi.org/10.1016/j.engstruct.2017.12.031>.
- Farrar CR, Doebling SW, Nix DA. Vibration-based structural damage identification. *Philos Trans.* 2001;359(1778):131–49. <https://doi.org/10.1098/rsta.2000.0717>.
- Taddei T, Penn J, Yano M, Patera A. Simulation-based classification; a model-order-reduction approach for structural health monitoring. *Arch Comput Methods Eng.* 2018;25(1):23–45.
- Doebling SW, Farrar C, Prime M. A summary review of vibration-based damage identification methods. *Shock Vibrat Digest.* 1998;30:91–105. <https://doi.org/10.1177/058310249803000201>.
- Farrar C, Worden K. *Structural health monitoring a machine learning perspective.* Hoboken: Wiley; 2013. <https://doi.org/10.1002/9781118443118>.
- Sohn H, Worden K, Farrar CR. Statistical damage classification under changing environmental and operational conditions. *J Intell Mater Syst Struct.* 2002;13(9):561–74. <https://doi.org/10.1106/104538902030904>.
- Entezami A, Shariatmadar H. Damage localization under ambient excitations and non-stationary vibration signals by a new hybrid algorithm for feature extraction and multivariate distance correlation methods. *Struct Health Monit.* 2019;18(2):347–75. <https://doi.org/10.1177/1475921718754372>.
- Eftekhar AS. *Online damage detection in structural systems.* Cham: Springer; 2014. <https://doi.org/10.1007/978-3-319-02559-9>.
- Bouzenad AE, Mountassir M, Yaacoubi S, Dahmene F, Koabaz M, Buchheit L, Ke W. A semi-supervised based k-means algorithm for optimal guided waves structural health monitoring: a case study. *Inventions.* 2019;4:1. <https://doi.org/10.3390/inventions4010017>.
- Entezami A, Shariatmadar H. An unsupervised learning approach by novel damage indices in structural health monitoring for damage localization and quantification. *Struct Health Monit.* 2018;17(2):325–45. <https://doi.org/10.1177/1475921717693572>.
- Goldstein M, Uchida S. A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PLOS ONE.* 2016;11(4):1–31. <https://doi.org/10.1371/journal.pone.0152173>.
- Bigoni C, Hesthaven JS. Simulation-based anomaly detection and damage localization: an application to structural health monitoring. *Comput Methods Appl Mech Eng.* 2020;363:112896. <https://doi.org/10.1016/j.cma.2020.112896>.
- Wang Z, Yan W, Oates T. Time series classification from scratch with deep neural networks: a strong baseline. In: *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 14–19 May, Anchorage, 2017. p. 1578–85. <https://doi.org/10.1109/IJCNN.2017.7966039>.
- Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science.* 2006;313(5786):504–7. <https://doi.org/10.1126/science.1127647>.
- Goodfellow I, Bengio Y, Courville A. *Deep Learning.* Boston: MIT Press; 2016. <http://www.deeplearningbook.org>.
- Pathirage CSN, Li J, Li L, Hao H, Liu W, Wang R. Development and application of a deep learning-based sparse autoencoder framework for structural damage identification. *Struct Health Monit.* 2019;18(1):103–22. <https://doi.org/10.1177/1475921718800363>.
- Choy WA. Structural health monitoring with deep learning. *Lecture Notes in Engineering and Computer Science.* In: *Proceedings of The International MultiConference of Engineers and Computer Scientists.* 2018. p. 557–60.
- Karim F, Majumdar S, Darabi H, Harford S. Multivariate LSTM-FCNs for time series classification. *Neural Netw.* 2019;116:237–45. <https://doi.org/10.1016/j.neunet.2019.04.014>.

20. Capellari G, Chatzi E, Mariani S. Structural health monitoring sensor network optimization through bayesian experimental design. *ASCE-ASME J Risk Uncertainty Eng Syst.* 2018;4:04018016. <https://doi.org/10.1061/AJRU6.0000966>.
21. Wang Q, Ripamonti N, Hesthaven JS. Recurrent neural network closure of parametric POD-Galerkin reduced-order models based on the Mori-Zwanzig formalism. *J Comput Phys.* 2020;410:109402.
22. Eftekhar Azam S, Bagherinia M, Mariani S. Stochastic system identification via particle and sigma-point kalman filtering. *Scientia Iranica.* 2012;19:982–91.
23. Teughels A, Maeck J, De Roeck G. Damage assessment by fe model updating using damage functions. *Comput Struct.* 2002;80:1869–79.
24. Entezami A, Shariatmadar H. Damage localization under ambient excitations and non-stationary vibration signals by a new hybrid algorithm for feature extraction and multivariate distance correlation methods. *Struct Health Monit.* 2019;18:347–75.
25. Eftekhar Azam S, Mariani S, Attari N. Online damage detection via a synergy of proper orthogonal decomposition and recursive bayesian filters. *Nonlinear Dyn.* 2017;89(2):1489–511.
26. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Levenberg J, Mané D, Monga R, Moore S, Murray D, Olah C, Schuster M, Shlens J, Steiner B, Sutskever I, Talwar K, Tucker P, Vanhoucke V, Vasudevan V, Viégas F, Vinyals O, Warden P, Wattenberg M, Wicke M, Yu Y, Zheng X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from [tensorflow.org](https://www.tensorflow.org/) 2015. <https://www.tensorflow.org/>.
27. Haykin S. *Neural networks and learning machines.* Upper Saddle River: Prentice Hall; 2009.
28. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *International Conference on Machine Learning (ICML)*, 6–11 July, Lille, France 2015.
29. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: *JMLR W&CP: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010)*, 13–15 May, vol. 9. Chia Laguna Resort, Sardinia, Italy, 2010. p. 249–56.
30. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *The IEEE conference on computer vision and pattern recognition (CVPR)*, 26 June–1 July, Boston, MA, 2015. p. 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>.
31. Kingma D, Ba J. Adam: A method for stochastic optimization. San Diego: University of Amsterdam; 2015. p. 1–13.
32. Karim F, Majumdar S, Darabi H. Insights into lstm fully convolutional networks for time series classification. *IEEE Access.* 2019;7:67718–25. <https://doi.org/10.1109/ACCESS.2019.2916828>.
33. De Callafon RA, Moaveni B, Conte JP, He X, Udd E. General realization algorithm for modal identification of linear dynamic systems. *J Eng Mech.* 2008;134(9):712–22. [https://doi.org/10.1061/\(ASCE\)0733-9399\(2008\)134:9\(712\)](https://doi.org/10.1061/(ASCE)0733-9399(2008)134:9(712)).
34. Corigliano A, Mariani S. Parameter identification in explicit structural dynamics: performance of the extended kalman filter. *Computer Methods Appl Mech Eng.* 2004;193(36–38):3807–35. <https://doi.org/10.1016/j.cma.2004.02.003>.
35. Bonnefoy-Claudet S, Cotton F, Bard P-Y. The nature of noise wavefield and its applications for site effects studies: a literature review. *Earth-Sci Rev.* 2006;79(3–4):205–27.
36. Ivanovic SS, Trifunac MD, Todorovska M. Ambient vibration tests of structures—a review. *ISET J Earthquake Technol.* 2000;37(4):165–97.
37. Capellari G, Chatzi E, Mariani S, Azam Eftekhar S. Optimal design of sensor networks for damage detection. *Procedia Eng.* 2017;199:1864–9.
38. Raudys SJ, Jain AK. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans Pattern Anal Mach Intell.* 1991;13(3):252–64.
39. Ribeiro RR, Lameiras RM. Evaluation of low-cost mems accelerometers for shm: frequency and damping identification of civil structures. *Latin Am J Solids Struct.* 2019. <https://doi.org/10.1590/1679-78255308>.
40. Ben-David S, Blitzer J, Crammer K, Kulesza A, Pereira F, Vaughan JW. A theory of learning from different domains. *Mach Learn.* 2010;79(1):151–75. <https://doi.org/10.1007/s10994-009-5152-4>.
41. Capellari G, Chatzi E, Mariani S et al. An optimal sensor placement method for shm based on bayesian experimental design and polynomial chaos expansion. In: *European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS)*, June 5–10, Athens, Greece, 2016. p. 6272–82.
42. Capellari G, Chatzi E, Mariani S. Cost-benefit optimization of structural health monitoring sensor networks. *Sensors.* 2018;18(7):2174. <https://doi.org/10.3390/s18072174>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.