applied sciences

MDPI

Article

An LSTM Autoencoder-Based Approach for Monitoring Railway Bridges

Viviana Giorgi [1,*], Ciro Tordela [2], Lorenzo Bernardini [1], Pablo Alex Ramírez Balbiano [3], Claudio Somaschini [1], Salvatore Strano [2] and Mario Terzo [2]

[1] Department of Mechanical Engineering, Politecnico di Milano, Via G. La Masa 1, 20156 Milan, Italy; lorenzo.bernardini@polimi.it (L.B.); claudio.somaschini@polimi.it (C.S.)
[2] Department of Industrial Engineering, University of Naples "Federico II", Via Claudio 21, 80125 Naples, Italy; ciro.tordela@unina.it (C.T.); salvatore.strano@unina.it (S.S.); mario.terzo@unina.it (M.T.)
[3] Displaid SRL Società Benefit, Via Carlo Freguglia 2, 20122 Milan, Italy; pablo.ramirez@displaid.it
* Correspondence: viviana.giorgi@polimi.it

Abstract

Continuous monitoring of railway bridges is essential for ensuring safety and operational reliability, considering aging mechanisms, rising traffic, and elevated speeds of railway vehicles. Frequently, traditional vibration-based approaches, including modal identification and data-driven diagnostic strategies, are strongly influenced by environmental and operational variability, requiring labeled damaged datasets or numerical simulations to provide reliable outcomes. However, the acquisition of complete and representative datasets for training neural networks in structural health monitoring remains a challenging task, particularly for large-scale civil structures such as bridges. In these cases, unsupervised learning approaches represent promising solutions. An unsupervised anomaly detection methodology for railway bridge monitoring based on a long short-term memory (LSTM) autoencoder (AE) trained exclusively on bridge accelerations under healthy structural conditions is proposed in the present work. Specifically, the acceleration responses are obtained from simulations made on a calibrated finite element model of the bridge, reproducing realistic train–bridge interaction scenarios. The multi-channel acceleration signals are reconstructed by the proposed LSTM AE to produce the Root Mean Square Error (RMSE) between measured and reconstructed acceleration responses as indicators of potential structural anomalies. A dual-threshold strategy is adopted for damage detection purposes, including a global threshold for identifying anomalies in the overall dynamic response and per-sensor thresholds derived from the healthy-condition RMSE distribution for detecting localized damages. Only healthy-condition data are required for employing the proposed technique, avoiding labeled damaged data for training purposes. The LSTM AE constitutes an effective and computationally efficient tool for anomaly detection and continuous structural health monitoring of railway bridges, as demonstrated by the obtained results, representing a promising alternative to classical modal-based approaches and existing deep learning-based methods.

Keywords: structural health monitoring; anomaly detection; long short-term memory; autoencoder; unsupervised learning; train–bridge interaction

Academic Editor: Michele D'Amato

Received: 24 March 2026
Revised: 17 April 2026
Accepted: 21 April 2026
Published: 27 April 2026

1. Introduction

Railway bridges are essential components of transportation networks; nowadays they are subjected to increasing traffic, growing dynamic loads and traveling speeds [1]. These

factors, combined with aging processes, threaten the structural integrity, safety and operability of the railway network. In this critical context, structural health monitoring (SHM) represents a fundamental tool for assessing structural conditions, planning maintenance activities and supporting life-cycle management strategies [2–6].

In the last few decades, SHM has gained growing interest in supporting and complementing traditional inspection procedures, typically based on visual surveys and periodic manual assessments, which are characterized by significant limitations and shortcomings [7], mainly related to their intrinsic subjectivity, inability to detect incipient or localized deterioration phenomena and the discretized frequency. Conversely, vibration-based SHM approaches potentially allow for quantitatively and continuously monitoring structural behavior. This can be achieved through the deployment of different sensing devices, such as accelerometers, velocimeters, linear variable displacement transducers (LVDTs), and fiber-optic sensors, among others. This, in practice, allows the extraction of dynamic (or static) indicators of structural integrity under in-service operational conditions [8–10].

Within vibration-based SHM, Operational Modal Analysis (OMA) has emerged as a central methodology for extracting modal parameters from structural response under operational and environmental actions [11]. Several studies have demonstrated the capability of OMA procedures, including covariance-driven stochastic subspace identification (SSI-COV), enhanced frequency domain decomposition (EFDD), transmissibility-based identification, and hybrid deterministic–stochastic techniques, to provide robust modal tracking even under challenging environmental and operational conditions [12,13]. These approaches have been successfully applied to steel truss bridges, reinforced concrete structures, arch bridges, and long-span cable-stayed systems, demonstrating effectiveness in long-term monitoring and in improving the interpretation of dynamic responses collected over extended time periods [14–20]. In fact, since modal parameters, i.e., natural frequencies and associated mode shapes and damping ratios, are related to structural condition [21,22], their tracking over time allows for an evaluation of the status of the monitored structure. However, it is well known that the natural frequencies of a bridge structure may be largely influenced by environmental factors, such as temperature and humidity [3,5,23]. Moreover, the reduced sensitivity to local damages may limit the early damage detection capability of modal-based indicators.

In recent years, machine learning (ML) and deep-learning techniques applied to SHM for railway bridges have been increasingly adopted to overcome the limitations of modal approaches [24,25]. Several works have explored drive-by (even called train-borne) detection methodologies, in which convolutional neural networks (CNNs) are trained on simulated or measured vehicle responses to identify structural damage indirectly through bogie accelerations [26]. Hybrid deep learning architectures, such as 1D-CNNs combined with Recurrent Neural Networks (RNNs), have been introduced in [21,26] to exploit synthetic datasets generated from calibrated finite element models to classify single or multiple damage scenarios. Additional developments include regression-based ML models for long-term monitoring using environmental and static measurements [23], digital-twin frameworks integrating vibration data with anomaly detection algorithms [27], and neural network-based methodologies relying on Gaussian processes, shallow ANNs, CNNs, or residual networks for predicting healthy accelerations and identifying deviations associated with structural deterioration [24,28–31].

Another form of proposed anomaly detection techniques is constituted of unsupervised ones. Algorithms such as one-class support vector machines, isolation forests, local outlier factor, and robust random cut forest have been adopted in several studies, demonstrating reliability in detecting abrupt changes through strain and acceleration features [32,33]. Other unsupervised approaches feature the use of statistical techniques such

as Extreme Value Theory [34], autoencoders [35,36], and DBSCAN paired with spectral clustering [37]. Several technical challenges still limit the effectiveness of existing approaches, although improvements have been made in the field of vibration-based structural health monitoring. Structural responses are strongly influenced by environmental and operational variability, which can mask or mimic damage-induced changes. Furthermore, many machine learning techniques are based on labeled datasets, including both healthy and damaged conditions, which are rarely available for real bridge structures. Finally, classical modal-based approaches exhibit limited sensitivity to localized damage, making early detection particularly challenging. Despite the previously mentioned promising results, most existing approaches exhibit limitations such as reliance on simulated datasets rather than long-term field measurements, and limited generalization capability to changing operational conditions. Unsupervised reconstruction-based strategies capable of learning the intrinsic dynamics of the healthy structure and detecting damage through deviations from the learned behavior have been explored in a limited number of works. These approaches are inherently well-suited for SHM applications, where structural responses are influenced by complex phenomena such as environmental fluctuations, train–bridge interaction, multi-modal vibration patterns, and sensor noise [32]. Moreover, given their nature, they overcome the limitations of requiring a comprehensive dataset referring to the damaged states of the structure.

The present work proposes an unsupervised SHM methodology for railway bridges based on a long short-term memory (LSTM) autoencoder (AE), trained exclusively on multi-sensor numerical acceleration time histories, referring to healthy structural conditions. Precisely, bridge accelerations are evaluated at the nodes of the FE model corresponding to the actual positions of the velocimeters installed on the actual structure.

The LSTM AE reconstructs the entire set of accelerations by exploiting temporal dependencies and spatial correlations related to bridge structural dynamic behavior. Deviations between measured and reconstructed signals are quantified through the root mean square error (RMSE), computed on ten-second time windows, enabling the identification of anomalous behaviors, a potential indication of the occurrence of structural damage.

Two complementary strategies were developed to enhance detection performance. A global window-level threshold, based on the distribution of healthy-condition reconstruction errors, is adopted to identify anomalies affecting the overall dynamic response. Moreover, per-sensor thresholds, defined according to the classical three-sigma statistical rule [38,39], are applied to the RMSE distribution concerning each sensor, enhancing sensitivity to localized damage. The proposed framework combines the advantages of deep learning, reconstruction-based anomaly detection, and multi-sensor response analysis, providing a data-driven, yet physically consistent tool for continuous monitoring of railway bridges under realistic operating scenarios simulated through a calibrated numerical model.

The main contributions of the present work can be summarized as follows:

- An unsupervised SHM methodology based on an LSTM-based autoencoder, trained exclusively on healthy-condition acceleration data, capable of detecting both global and localized structural anomalies without requiring labeled damaged datasets;
- A reconstruction-based damage indicator, defined through the RMSE, which allows for detecting subtle variations in the structural behavior that may not be reflected in modal parameter changes or frequency-based metrics;
- A dual-threshold anomaly detection strategy, including a global window-level threshold and sensor-specific thresholds derived from healthy-condition error distributions, which enhances sensitivity to localized deterioration while preserving robustness against environmental and operational variability;

- A methodology providing a complementary data-driven framework for attempting to mitigate limitations of OMA and supervised ML approaches, particularly in scenarios where labeled damage data are not available.

To better position the proposed methodology within the current literature, a structured comparison with recent deep learning-based approaches for structural health monitoring (SHM) is provided. While studies based on LSTM, autoencoder (AE), and graph-based architectures have shown promising capabilities for damage detection, significant differences remain regarding data requirements, model complexity, and evaluation settings.

Specifically, LSTM-based approaches, such as [40], typically rely on supervised learning and require labeled damage datasets, often focusing on indirect measurements derived from vehicle–bridge interactions. In contrast, the framework presented here is fully unsupervised and utilizes direct multichannel bridge accelerations. Regarding autoencoder-based methods, works such as [41] employ reconstruction-based anomaly detection using vehicle-based responses in a drive-by setup, which offers limited spatial exploitation; our approach instead utilizes direct monitoring to enable explicit spatial-temporal modeling. While ref. [42] explores comparative AE methods for general SHM, the present study is specifically tailored to railway bridges by leveraging LSTM-driven temporal modeling. Furthermore, although the method in [43] provides near real-time anomaly detection using numerical and experimental data on generic structures, it lacks a refined validation scheme. The current framework addresses this by incorporating a dual-threshold approach—combining global and per-sensor thresholds—to improve detection sensitivity and localization. Finally, while graph-based methods like those in [44] can explicitly model spatial relations via Graph Neural Networks (GNNs), they generally necessitate high architectural complexity and large, structured datasets. Compared to these approaches, the proposed methodology adopts a more streamlined LSTM-autoencoder architecture. It remains fully unsupervised, training exclusively on healthy-condition accelerations, and effectively exploits multichannel temporal data without the computational overhead of graph-based modeling.

The remaining part of the paper is organized as follows: Section 2 describes the case study under analysis. Subsequently, in Section 3, the attention is focused on the description of the finite element model of the bridge adopted in this work. The AI-based framework, developed for anomaly detection purposes, is described in Section 4. Section 5 shows and describes the results obtained from simulations, analyzing the performance of the proposed methodology. Finally, conclusions and future perspectives are discussed in Section 6.

2. Case Study

The bridge under study spans the Ticino River, located in the municipality of Turbigo, between the cities of Turbigo and Galliate in northern Italy. It was built in 1952 to replace an older structure, damaged after World War II.

The bridge, shown in Figure 1, features a steel truss girder design with four supports, all made of reinforced concrete. The lateral supports rest on two box abutments located on both sides of the river, while the central supports are positioned on two piers within the river. The structure has two decks: the upper deck serves as a railway line (single track), made entirely of steel. Instead, the lower deck is designated for road vehicles (with two lanes) and includes two lateral steel grillages for pedestrian use. The bridge span is divided into three sections: 44 m on Galliate's side, 56 m for the mid-span, and 45 m on Turbigo's side. Additionally, the bridge incorporates two Gerber connections in the mid-span, both located at one-fifth of the span length.

Figure 1. Side view of the actual bridge spanning Ticino River between the municipalities of Turbigo and Galliate.

The monitoring system installed on this railway bridge is composed, among the others, of twenty-two velocimeters (i.e., SARA SS45) dedicated to measuring vertical velocities along the vertical axis (z), promptly converted into accelerations. The sensors are distributed along the longitudinal axis (x) of the deck and arranged into two parallel rows corresponding to the upstream and downstream sides of the structure, as schematically depicted in Figure 2.

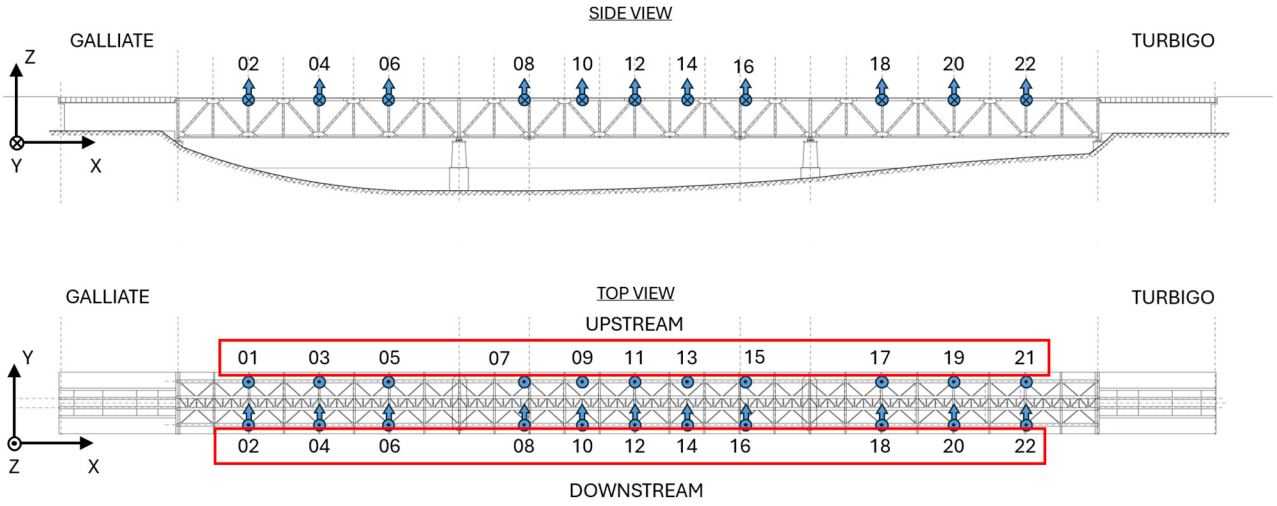

Figure 2. Sensors layout on the monitored bridge (side and top views are shown). Sensor IDs are highlighted in red.

Each sensor is mounted directly onto the main longitudinal upper girders, providing a homogeneous sampling of the vertical response along the span. The top view emphasizes the spatial organization of the sensing grid: odd-indexed sensors (IDs 01–21) are located on the upstream side of the deck, whereas even-indexed sensors (IDs 02–22) are aligned downstream. It is important to notice that only sensors installed on the downstream side of the bridge acquire data along the lateral direction (y-axis). A dense spatial discretization of the vertical dynamic response is ensured by the presented configuration, enhancing the monitoring system to detect both global and localized changes in structural behavior, including potential asymmetries between the two sides of the deck.

The complete sensor array potentially enables the proposed Artificial Intelligence (AI)-based methodology for anomaly detection tasks. The multi-sensor layout, combined with its symmetric deployment, provides a reliable basis for identifying deviations from the healthy structural state. In this paper, numerical accelerations refer to the nodes located in correspondence with the sensors mounted on the actual structure.

3. Models

To accurately simulate the response of the bridge, a calibrated FE model was designed, shown in Figure 3a. This step is crucial, as the accuracy of subsequent damage simulations depends on the reliability and representativeness of the model. The modeling process consisted of two steps: an initial modeling phase, during which the primary assumptions and structural properties are defined, and a calibration phase, where the model parameters are refined and tuned to closely match the actual behavior of the bridge.

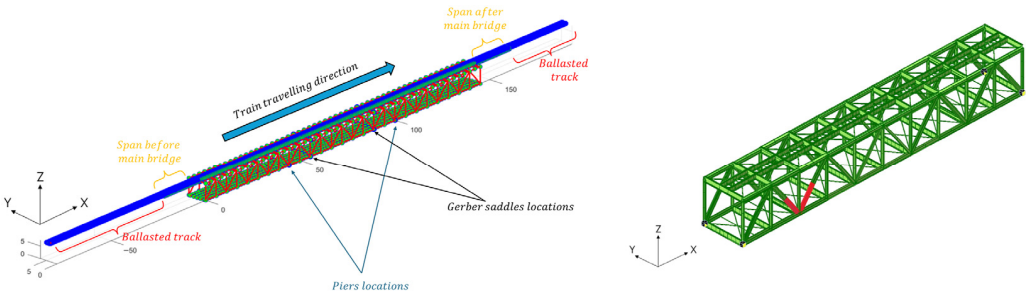

Figure 3. (**a**) Finite element model of the monitored bridge. Before and after the bridge two ballasted track sections are inserted to avoid fictitious transients due to rail vehicle motion along the structure; (**b**) central span of the bridge. Damaged elements are drawn in red (for LD10 and LD70 scenarios).

Two structural analysis software were used in this work, i.e., Midas GEN 2025 v1.2 and ADTreS [45]. The first is used to create the initial model of the bridge, incorporating its geometry and structural properties. This stage relied on technical drawings created as part of the most recent studies and subsequent interventions. Later, the model was translated into the input file format for ADTreS, a non-commercial software, developed over the years in the Department of Mechanical Engineering of Politecnico di Milano. This software allows for time-domain simulation of the dynamic interaction between bridge, track, and train, as illustrated in detail in [45–47].

Two types of FEM elements were used: the Euler–Bernoulli beam (EB) for steel trusses and the Kirchhoff–Love plate (KL) for the roadway deck. A total of 32 cross-sections, taken from the technical drawings of the structure, were identified for the different beam elements. A constant thickness was assumed for the reinforced concrete slab (equal to 23 cm). The Gerber connections, positioned in the central span of the structure, were modeled using 3D visco-elastic elements, whose properties are collected in Table A2. Main mechanical and geometrical properties are collected in Appendix A, in Tables A1 and A2.

Subsequently, model calibration involved the comparison of bridge numerical properties with those obtained by measurements taken from the actual structure during a preliminary testing campaign. An OMA approach was used to extract bridge modal parameters, whose algorithm is described in detail in [2]. The obtained natural frequencies of the bridge FE model are gathered in Table 1, compared with experimental counterparts computed during the preliminary campaign. By computing the MAC between numerical and experimental mode shapes, it is possible to notice that vertical mode shapes are better represented by the calibrated model compared to the lateral ones. However, the frequencies associated with the two lateral modes present differences below the 2% with respect to experimental outcomes.

Table 1. Comparison between numerical and experimental mode shapes for the bridge under analysis.

Mode	Experimental [Hz]	Numerical [Hz]	Error [%]	MAC
1st Vertical	3.04	3.15	3.62	0.99
1st Lateral	3.38	3.42	1.18	0.84
2nd Lateral	3.42	3.35	2.05	0.82
2nd Vertical	4.10	4.15	1.22	0.99
3rd Vertical	4.66	4.50	3.43	0.98

The train–bridge interaction simulations were carried out by considering multiple train speeds and loading configurations to investigate how changes in the passing vehicle affect the structural response and test the robustness of the algorithm. The adopted vehicle is representative of a regional commuter train, daily transiting over the monitored bridge. It consists of four coaches, each supported by two bogies; in turn, each bogie comprises two axles. Each rail coach is described by 37 degrees of freedom, and it is assumed to move at constant forward velocity. As mentioned, a set of traveling speeds was tested, from 40 to 60 km/h, with a spacing of 2 km/h. Moreover, different coach masses were considered in an attempt to represent different passengers' occupation of the train during the daily traffic. Being this work a preliminary study to investigate the capability of the presented algorithm in capturing structural damages, track irregularity is not accounted for and measuring noises are neglected. However, a certain degree of variability in operational loading scenarios is considered in this work. Precisely, a passenger train with five different configurations in terms of car body mass is modeled, traveling on the structure with different speeds, for different runs (see Section 4.1 for details).

As briefly recalled before, train–track–bridge dynamic simulation was carried out with ADTreS software [45], which enables FE modeling of track/bridge structures and the modeling of the rail vehicle through a multi-body approach. Structure and train are represented as two different sub-systems coupled through contact-forces, exchanged at the wheel–rail interface [46]. Time integration is performed using a modified Newmark approach (to account for non-linearities in the problem), discussed by Bernardini et al. [48].

Regarding damages applied to the numerical model of the bridge under analysis, three scenarios are considered:

1. Global corrosion (GC) implemented through a uniform reduction of 5% in both Young's modulus and material density (for steel material) applied to all truss elements of the structure. The test enables the assessment of whether the anomaly indicators can detect changes that manifest coherently across all sensing locations, considering a spatially distributed mild degradation.
2. Localized damage (LD) in which a localized damage in the proximity of sensor 8 (see Figure 2) is implemented through reductions of 10% in elastic modulus and density in a lateral joint of the central span. In the remaining part of the paper, this scenario is referred to as LD10.
3. Another LD case (i.e., LD70) in which a localized damage in the proximity of sensor 8 is implemented through reductions of 70% in elastic modulus and density.

The elements involved in the LD10 and LD70 damage cases are pictured in Figure 3b.

The objective of the first scenario is to assess whether the anomaly indicators can detect changes that manifest coherently across all sensing locations, considering a spatially distributed mild degradation. Instead, the third and fourth scenarios emulate progressively severe local defects, allowing for evaluating whether the reconstruction error produced by the LSTM AE increases specifically in correspondence with the affected sensors while remaining consistent with healthy behavior elsewhere. Particularly, the LD70 case, introduces a significant stiffness reduction and is expected to induce more evident distortions in

the vibration patterns captured by the affected channels (i.e., accelerations evaluated at the node close to the localized damage).

4. General Framework of the LSTM AE-Based Monitoring Technique

The proposed monitoring framework processes the twenty-two vertical acceleration signals collected along the bridge at a sampling frequency of $f_s = 500$ Hz. The continuous data stream acquired from the sensing system is segmented into fixed-length time windows of 10 s, corresponding to $N_W = 5000$ samples per sensor. The multichannel acceleration signal is defined as:

$$\boldsymbol{A}(t) = [a_{z,1}(t) \;\; a_{z,2}(t) \;\; \ldots \;\; a_{z,22}(t)] \tag{1}$$

with $\mathbf{A}(t) \in \mathbb{R}^{1\times 22}$. The signal is partitioned using a sliding-window scheme with partial overlap. By denoting with L the window length and S the step size ($S < L$), the i-th window $\mathbf{X}_i$ is defined as:

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{A}(t_i) \\ \mathbf{A}(t_i+1) \\ \vdots \\ \mathbf{A}(t_i+L-1) \end{bmatrix} \in \mathbb{R}^{L\times 22} \tag{2}$$

where the window start index is given by:

$$t_i = 1 + (i-1)S \tag{3}$$

where the subscript $i = 1, \ldots, n_W$ denotes the index of the sliding window in the segmentation sequence.

In the present configuration, $L = 5000$ samples and $S = 4000$ samples, yielding a two-second overlap between consecutive windows. Therefore, the full ten-second segment of all twenty-two acceleration channels is contained in each window $\mathbf{X}_i$.

The LSTM AE is independently fed by each multichannel window $\mathbf{X}_i$ for encoding the temporal dynamics of the multivariate signal and reconstructing a corresponding output window $\hat{\mathbf{X}}_i$.

The deviation between the measured and reconstructed vertical accelerations is quantified through the reconstruction error matrix:

$$\mathbf{E}_i = \mathbf{X}_i - \hat{\mathbf{X}}_i \tag{4}$$

The per-sensor Root Mean Square Error (RMSE) is computed as:

$$\mathrm{RMSE}_{i,j} = \sqrt{\frac{1}{L}\sum_{k=1}^{L}\left(X_i(k,j) - \hat{X}_i(k,j)\right)^2}, j = 1, \ldots, 22 \tag{5}$$

Therefore, for each time window, a feature vector $\mathrm{RMSE}_i \in \mathbb{R}^{1\times 22}$ that characterizes the reconstruction quality of the full sensor array is obtained.

The RMSE values are subsequently compared against sensor-level (local) and global thresholds, learned from healthy-condition data during the training stage of the LSTM AE, to determine whether window $\mathbf{X}_i$ and its associated sensors are consistent with nominal structural behavior or indicative of a potential anomaly. The entire procedure is summarized schematically in Figure 4. The proposed approach introduces specific design choices tailored to multichannel structural monitoring of railway bridges, extending standard reconstruction-based anomaly detection approaches. Multichannel vibration data are exploited to jointly capture temporal and spatial dependencies of the structural re-

sponse, rather than analyzing each sensor independently. Furthermore, anomaly detection is carried out by introducing a dual-threshold strategy, combining a global window-level criterion with sensor-specific thresholds, enabling the identification of both distributed and localized anomalies. Finally, the methodology adopts a fully unsupervised training strategy based exclusively on healthy-condition datasets, making it suitable for SHM applications where labeled damage data are not available.

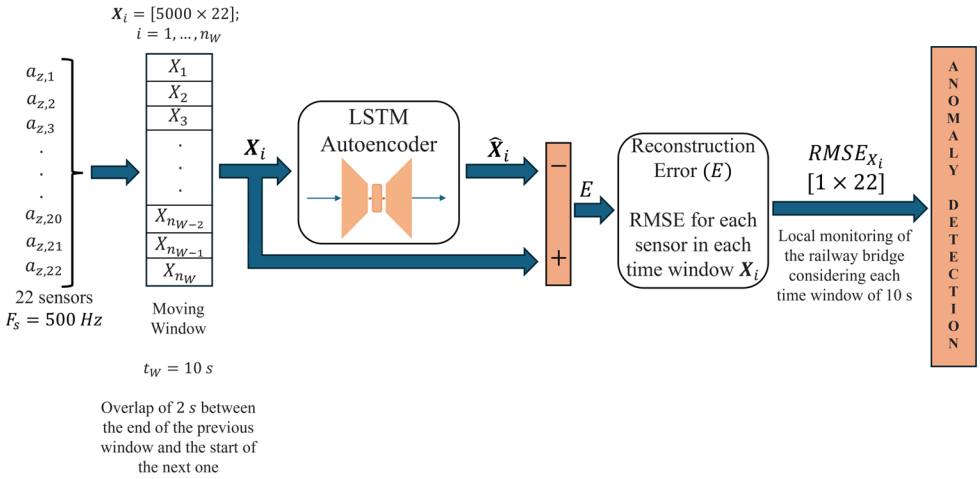

Figure 4. Workflow of the proposed LSTM AE-based monitoring technique.

The proposed neural network architecture is designed as a sequence-to-sequence LSTM AE [49,50] specifically developed for the reconstruction of multichannel vibration signals associated with the dynamic behavior of the monitored railway bridge. The neural network model operates on time series composed of twenty-two simultaneous acceleration channels, exploiting both the temporal dependencies and the spatial correlations inherent in the structural response. By learning a compact latent representation of the normal-state dynamics and by enforcing accurate sequence reconstruction, the LSTM AE provides a reliable, fully data-driven basis for identifying deviations indicative of structural anomalies. The LSTM AE is composed of three main components: the encoder, the bottleneck, and the decoder, represented by three LSTM layers, respectively. For a more detailed description of the architecture of the LSTM AE, please refer to Appendix B.1.

4.1. Training Dataset Generation

The training dataset used to calibrate the LSTM AE is constructed exclusively from the healthy structural conditions of the monitored railway bridge. To this end, a comprehensive numerical campaign is performed by simulating the coupled train–bridge dynamic interaction using the modeling framework described in Section 2. Five railway vehicle configurations, denoted as Train-A, Train-B, Train-C, Train-D and Train-E, are considered. Specifically, each railway vehicle consists of four car bodies, each car body is mounted on two bogies, and each bogie is equipped with two wheelsets. The properties of the five configurations differ only in the mass of the four car bodies, as reported in Table 2, introducing variability in the dynamic excitation.

The dynamic railway vehicle–bridge interaction is simulated over a range of cruising speeds from 40 km/h to 60 km/h, with increments of 2 km/h, for each railway vehicle. The vertical acceleration response of the bridge is recorded during each simulated railway vehicle passage, at the twenty-two positions corresponding to the ones related to the virtually mounted sensors on the bridge model, sampled at $f_s = 500$ Hz. By considering a rail-

way vehicle configuration $T \in \{\text{Train-A, Train-B, Train-C, Train-D, Train-E}\}$, the resulting multi-sensor signal is denoted as:

$$\boldsymbol{X}_{\text{healthy},\boldsymbol{T}} = \boldsymbol{a}_{\boldsymbol{z},\boldsymbol{T}} \in \mathbb{R}^{L_T \times 22} \tag{6}$$

where L_T is the total number of samples obtained by concatenating the acceleration histories corresponding to all simulated speeds. An example of the resulting multi-speed time series for Sensor 11 considering Train-A is shown in Figure 5, illustrating the succession of vibration signatures generated by the speed-incremented passages.

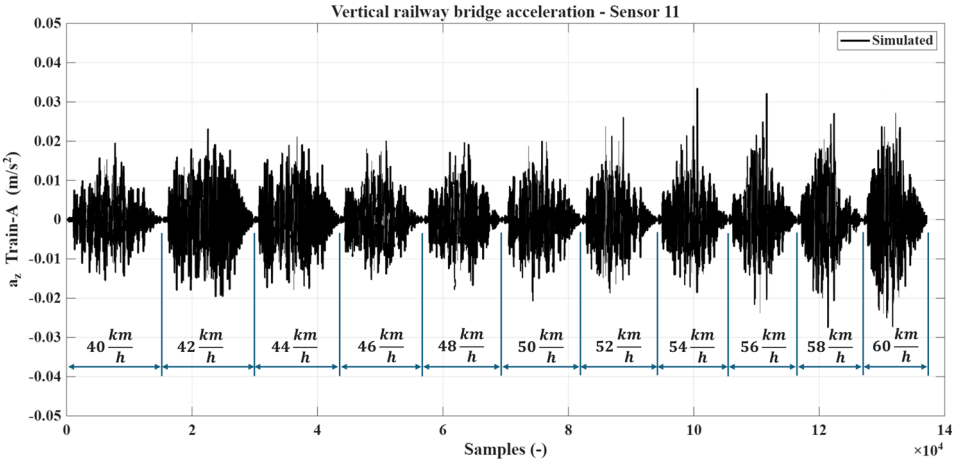

Figure 5. Simulated and concatenated vertical accelerations on Sensor 11.

Table 2. The main characteristics of the considered railway vehicles. Variability is introduced in the car body mass to simulate different passenger occupations among different runs. Please notice that stiffness and damping values for primary and secondary suspensions refer to each single bogie.

	Train-A	Train-B	Train-C	Train-D	Train-E
Car body 1 mass [kg]	48,020	46,372	49,310	49,307	44,470
Car body 2 mass [kg]	45,852	38,996	46,353	42,286	45,936
Car body 3 mass [kg]	39,246	40,531	39,506	44,960	44,890
Car body 4 mass [kg]	48,911	45,600	49,428	41,940	49,330
Bogie 1 mass [kg]	3318	3318	3318	3318	3318
Bogie 2 mass [kg]	6100	6100	6100	6100	6100
Wheelset 1 mass [kg]	1916	1916	1916	1916	1916
Wheelset 2 mass [kg]	1926	1926	1926	1926	1926
Prim. vert. stiff. 1 [kN/m]	3832	3832	3832	3832	3832
Prim. vert. damp. 1 [kNs/m]	40	40	40	40	40
Prim. vert. stiff. 2 [kN/m]	5200	5200	5200	5200	5200
Prim. vert. damp. 2 [kNs/m]	64	64	64	64	64
Sec. vert. stiff. 1 [kN/m]	770	770	770	770	770
Sec. vert. damp. 1 [kNs/m]	72	72	72	72	72
Sec. vert. stiff. 2 [kN/m]	830	830	830	830	830
Sec. vert. damp. 2 [kNs/m]	72	72	72	72	72

The fixed-size input sequences for the LSTM AE are obtained by segmenting each multichannel signal $X_{\text{healthy},T}$ into partially overlapping windows of 10 s duration. For each train configuration T, the i-th window is defined as:

$$W_{\text{healthy},T,i} = \begin{bmatrix} X_{\text{healthy},T}(t_i, 22) \\ X_{\text{healthy},T}(t_i + 1, 22) \\ \vdots \\ X_{\text{healthy},T}(t_i + L - 1, 22) \end{bmatrix} \in \mathbb{R}^{L \times 22} \tag{7}$$

where the window-start index is:

$$t_i = 1 + (i-1)S, \;\; i = 1, 2, \ldots, N_T \tag{8}$$

where $L = 5000$ and $S = 4000$ denote the number of samples per window (10 s at 500 Hz) and the step size between consecutive windows, respectively, corresponding to an overlap of 2 s, and N_T denotes the number of 10 s segments, which corresponds to 34 time windows per train. The previously described window length was adopted as a compromise between temporal resolution and the need to preserve a sufficiently rich description of the bridge's vibration response during train passages. The LSTM AE is enabled to capture multiple cycles of the dominant bending and torsional modes, as well as the spatially evolving excitation produced by successive axle crossings. Shorter windows would have reduced the amount of dynamic information available to the network, limiting its ability to learn medium- and long-range temporal dependencies and to represent non-stationary effects induced by the moving load. Longer time windows have not been chosen due to the higher computational and memory requirements of the sequence-to-sequence training procedure without providing commensurate benefits in terms of anomaly detectability. The adopted duration, therefore, offers a balanced trade-off between dynamic content, computational efficiency, and temporal resolution.

The total number of windows across the five train configurations is:

$$N_{\text{segments}} = \sum_{T} N_T = 170 \tag{9}$$

Each of the collected N_{segments} represents a ten-second realization of the bridge response under healthy operating conditions. All time windows are then aggregated into a single dataset:

$$W_{\text{healthy,TOT}} = \left\{ W_{\text{healthy},T,i} \;\middle|\; T \in \{\text{Train-A, Train-B, Train-C, Train-D, Train-E}\},\; i = 1, \ldots, N_T \right\} \tag{10}$$

which constitutes the complete collection of healthy-condition samples. The healthy dataset is divided into two subsets to avoid data leakage. Specifically, 70% of $W_{\text{healthy,TOT}}$ named W_{healthy}, equalling 119 time windows, is employed for training the LSTM AE. The remaining 30% of $W_{\text{healthy,TOT}}$, equal to 51 time windows, is used for testing the proposed LSTM AE. Furthermore, the testing healthy dataset is employed to define both the global and sensor-specific anomaly detection thresholds, described in the following Sections, based on the reconstruction error. A standardization process, based on the Z-score rule, is made on each time window before feeding it to the LSTM AE. For each sensor channel j, the normalized sample is computed as:

$$W_{\text{healthy,std}}(k, j) = \frac{W_{\text{healthy}}(k, j) - \mu_{healthy,j}}{\sigma_{healthy,j}}, \quad k = 1, \ldots, 5000 \tag{11}$$

where $\mu_{healthy,j}$ and $\sigma_{healthy,j}$ denote the sample mean and standard deviation of channel $j = 1, \ldots, 22$, respectively, computed across the entire healthy dataset. Furthermore, the standardization parameters used in the Z-score consist of two vectors $\boldsymbol{\mu_{healthy}}$, $\boldsymbol{\sigma_{healthy}} \in \mathbb{R}^{1 \times 22}$ each containing the sample mean and standard deviation computed in-

dependently for each of the twenty-two sensor channels, respectively. This channel-wise standardization ensures a consistent scaling of the input features, preserving the relative dynamic variability across the sensing array, which is essential for the stable training of the LSTM AE.

The LSTM AE presented in Section 4, and detailed in Appendix B.1, is trained employing the final training dataset $W_{\text{healthy,std-perm}}$, which is constituted of randomly permuted standardized time windows. The random permutation removes any residual ordering effects associated with railway vehicle type or speed. A schematic representation of the dataset-generation workflow is provided in Figure 6.

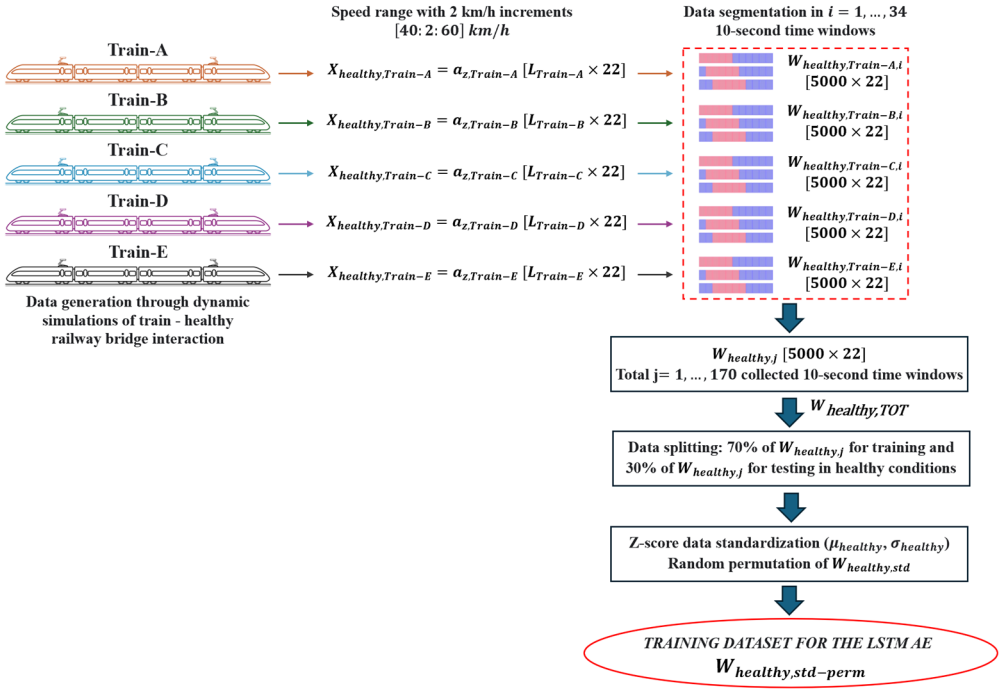

Figure 6. Workflow of the training dataset generation.

4.2. LSTM AE Training and Optimization

The LSTM AE training process is integrated into a Bayesian Optimization (BO) framework for identifying the configuration of hyperparameters that minimizes the reconstruction error. Furthermore, the coupling between the BO and the LSTM AE training ensures stable convergence and adequate regularization under the heterogeneous excitation scenarios considered in the healthy-condition dataset.

Details on the theory behind Bayesian Optimization and how it was performed are included in Appendix B.2.

The optimization converges towards a configuration that achieves a favorable trade-off between network expressiveness, generalization ability, and training stability.

The selected configuration features include an encoder of 116 units, a bottleneck layer of 58 units defined as half of the encoder size, and a decoder symmetric to the encoder.

The values of optimized dropout probabilities associated with the encoder and bottleneck outputs are 0.15545 and 0.20608, respectively, indicating that moderate stochastic regularization was instrumental in preventing co-adaptation of recurrent features. The initial learning rate converges to 0.0018109, ensuring both sufficiently rapid initial descent and stable refinement of the weights near convergence. The optimized ℓ_2-regularization coefficient assumes the value 1.1654×10^{-6}, effectively limiting weight growth without suppressing the capability of the LSTM AE to represent the intrinsic variability of healthy-condition vibrations. An optimal mini-batch size of 16 samples is selected, reflecting the

gradient sensitivity to temporal fluctuations in multichannel sequences. A summary of the optimized hyperparameters and the associated training performance is reported in Table 3.

Table 3. Optimized hyperparameters and resulting training performance.

Hyperparameter	Value
Encoder units	116
Bottleneck units	58
Decoder units	116
Dropout 1	0.15545
Dropout 2	0.20608
Initial learning rate	0.0018109
ℓ_2-regularization	1.1654×10^{-6}
Mini-batch size	16
Final RMSE (Bayesian Optimization)	0.046981
Final training MSE	0.014
Epochs	100

A quick decrease in the RMSE is observable in Figure 7 during the initial iterations of the BO. The surrogate model progressively identifies hyperparameter regions associated with improved reconstruction performance. A plateau is reached by the BO procedure around the tenth iteration, indicating that the search space contains a relatively narrow region of high-performing configurations. The early stabilization of the RMSE also confirms the adequacy of the selected hyperparameter ranges, which are sufficiently broad to encompass the optimal configuration yet restrictive enough to avoid pathological regimes that could lead to unstable training dynamics.

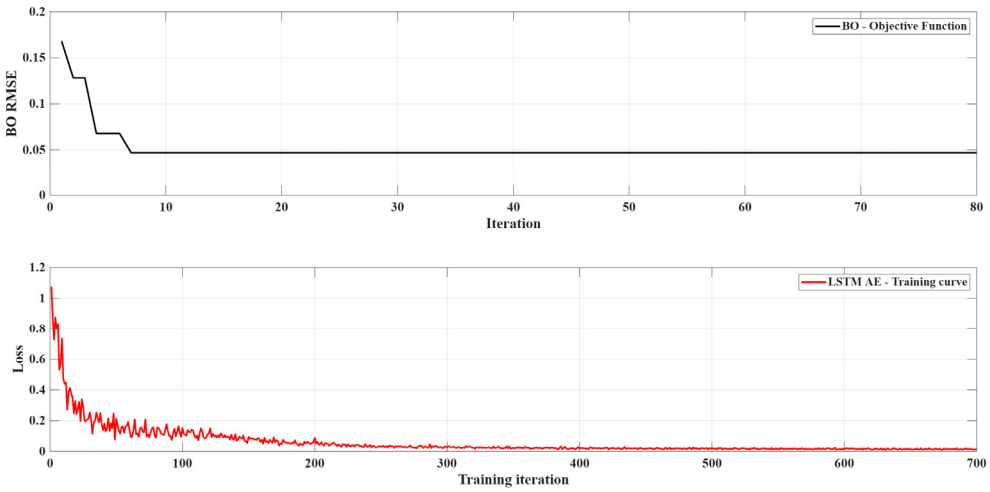

Figure 7. Bayesian Optimization objective function and optimized LSTM AE training curves.

Furthermore, Figure 7 shows the training curve of the optimal LSTM AE, which exhibits a monotonic decay of the loss function, confirming that the network learns to reproduce the healthy-condition vibration patterns. Specifically, a sharp decrease in the loss function can be observed in the first 100–150 iterations, reflecting the rapid adjustment of the recurrent and fully connected layers in capturing the dominant components of the input sequences.

A steady refinement phase over several hundred iterations follows the initial transient, in which the network progressively stabilizes its internal representation and reduces residual reconstruction errors.

The absence of oscillations, divergences, or sudden loss spikes confirms the numerical stability of the optimized hyperparameters and demonstrates the effectiveness of the adopted regularization strategy, particularly the moderate dropout rates and ℓ_2-penalty, in preventing overfitting.

The complete training process, integrated with the Bayesian Optimization one, required approximately one day on a workstation equipped with an NVIDIA GeForce RTX 3090 GPU. Although the number of independent simulations is limited, the dataset comprises many multichannel time windows obtained through signal segmentation. A sufficiently rich representation of the structural response under different operating conditions is provided by the used windows, including variations in train speed and loading configurations. Furthermore, the proposed approach is fully unsupervised, aiming to learn the nominal structural behavior in healthy conditions rather than performing classification tasks, which reduces the dependency on large, labeled datasets.

4.3. Anomaly Detection Criteria

After training the LSTM AE on healthy-condition data, anomaly detection is performable by evaluating the discrepancy between the measured multichannel acceleration windows and their reconstructed counterparts. Since the training dataset was standardized using the channel-wise mean and standard deviation vectors $\boldsymbol{\mu_{healthy}}$, $\boldsymbol{\sigma_{healthy}} \in \mathbb{R}^{1\times 22}$, stored during the dataset generation described in Section 4.1, all reconstructed windows are first de-standardized using these parameters before computing the detection metrics, ensuring that reconstruction errors are assessed in physically meaningful units consistent across both training and testing phases.

For a given window $W_i \in \mathbb{R}^{L\times 22}$ and sensor channel j, the reconstruction error is quantified by the Root Mean Square Error (RMSE):

$$\mathrm{RMSE}_{i,j} = \sqrt{\frac{1}{L}\sum_{k=1}^{L}\left(W_i(k,j) - \hat{W}_i(k,j)\right)^2} \tag{12}$$

where $L = 5000$ is the window length in samples. The set of per-sensor reconstruction errors for window i forms the vector:

$$\mathrm{RMSE}_i = [\mathrm{RMSE}_{i,1}, \ldots, \mathrm{RMSE}_{i,22}] \in \mathbb{R}^{1\times 22} \tag{13}$$

A global anomaly indicator for each window is defined as the Euclidean norm of this vector:

$$E_i = \| \mathrm{RMSE}_i \|_2 = \sqrt{\sum_{j=1}^{22} \mathrm{RMSE}_{i,j}^2} \tag{14}$$

Statistic-based thresholds for anomaly detection [51,52] are computed exclusively from healthy-condition training windows, and once calculated, they are stored for use in operational monitoring. For each sensor j, the threshold, based on the three-sigma statistical rule, is defined as:

$$\mathrm{thr}_j = \mu_j^{(\mathrm{RMSE})} + 3\sigma_j^{(\mathrm{RMSE})} \tag{15}$$

where $\mu_j^{(\mathrm{RMSE})}$ and $\sigma_j^{(\mathrm{RMSE})}$ are the mean and standard deviation of $\mathrm{RMSE}_{i,j}$ computed across all healthy training windows. Analogously, the global threshold associated with the indicator E_i is given by:

$$\mathrm{thr}_E = \mu_E + 3\,\sigma_E, \quad \mu_E = \mathrm{mean}(E_i),\ \sigma_E = \mathrm{std}(E_i) \tag{16}$$

A window is flagged as anomalous whenever either the sensor-level error or the global indicator exceeds the corresponding healthy threshold. The three-sigma rule, based on the empirical distribution of the reconstruction errors, has been chosen to provide a simple and interpretable statistical criterion for identifying deviations from nominal behavior observed under healthy conditions linked to the testing dataset described in Section 4.1. Furthermore, the 3σ-based threshold has been selected, after a sensitivity analysis, as a compromise between the 2σ- and 4σ-based thresholds. Particularly, the lower threshold 2σ led to excessive sensitivity, resulting in false positives in both healthy and faulty conditions. In contrast, the higher threshold 4σ reduced the sensitivity of the proposed method, limiting its capability to detect damage scenarios. Therefore, the 3σ-based threshold represents a trade-off between reliability and sensitivity, minimizing false alarms during anomaly detection.

For sensor j, the Positive Rate (PR) is therefore computed as:

$$\mathrm{PR}_j = \frac{1}{N_{\mathrm{win}}} \sum_{i=1}^{N_{\mathrm{win}}} \mathbb{I}\left(\mathrm{RMSE}_{i,j} > \mathrm{thr}_j\right) \tag{17}$$

where N_{win} is the number of evaluated windows and $\mathbb{I}(\cdot)$ is the indicator function:

$$\mathbb{I}(A) = \begin{cases} 1, & if\ the\ event\ A\ is\ true \\ 0, & if\ the\ event\ A\ is\ false \end{cases} \tag{18}$$

The corresponding complement is defined as the Reliability Index (RI) of the j-th sensor:

$$\mathrm{RI}_j = 1 - \mathrm{PR}_j \tag{19}$$

Similarly, the global Positive Rate based on the composite indicator E_i is:

$$\mathrm{PR}_{\mathrm{glob}} = \frac{1}{N_{\mathrm{win}}} \sum_{i=1}^{N_{\mathrm{win}}} \mathbb{I}(E_i > \mathrm{thr}_E) \tag{20}$$

with global RI:

$$\mathrm{RI}_{\mathrm{glob}} = 1 - \mathrm{PR}_{\mathrm{glob}} \tag{21}$$

The reliability of the detection process has been enhanced by introducing a stricter two-window consistency rule. Under this criterion, an anomaly is declared only if two consecutive windows exceed the same threshold. For sensor j, this corresponds to:

$$\mathbb{I}\left(\mathrm{RMSE}_{i,j} > \mathrm{thr}_j\right) \wedge \mathbb{I}\left(\mathrm{RMSE}_{i+1,j} > \mathrm{thr}_j\right) \tag{22}$$

where $\wedge$ denotes the logical AND operator. The associated positive rate is therefore:

$$\mathrm{PR}_j^{(2\mathrm{win})} = \frac{1}{N_{\mathrm{win}}} \sum_{i=1}^{N_{\mathrm{win}}} \mathbb{I}\left(\mathrm{RMSE}_{i,j} > \mathrm{thr}_j \wedge \mathrm{RMSE}_{i+1,j} > \mathrm{thr}_j\right) \tag{23}$$

with RI:

$$\mathrm{RI}_j^{(2\mathrm{win})} = 1 - \mathrm{PR}_j^{(2\mathrm{win})} \tag{24}$$

An analogous formulation applies to the global indicator E_i under the same consistency rule. The complete anomaly detection framework, which includes both the single-window and the two-window criteria, establishes a set of decision rules that can later be applied to identify deviations from the healthy structural behavior, including potential cases where consecutive windows exhibit consistent increases in reconstruction error. Additionally, the

Precision, Recall, and the F1-score are computed to highlight the detection performance across the considered scenarios. Specifically, by considering the following ground-truth:

$$y_{truth,j} = \begin{cases} 1, & if\ damaged \\ 0, & if\ healthy \end{cases};\ j = 1, \ldots, 22 \text{ sensors} \tag{25}$$

a candidate sensor $y_j (j = 1, \ldots, 22)$ takes the value 1, detecting damage, if and only if $PR_j > 0$, otherwise y_j takes the value 0, highlighting the healthy condition of the bridge around its positioning. Therefore, the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values can be computed as follows:

$$\begin{aligned} TP_j &= (y_j = 1) \wedge \left(y_{truth,j} = 1\right) \\ FP_j &= (y_j = 1) \wedge \left(y_{truth,j} = 0\right) \\ TN_j &= (y_j = 0) \wedge \left(y_{truth,j} = 0\right) \\ FN_j &= (y_j = 0) \wedge \left(y_{truth,j} = 1\right) \end{aligned} \tag{26}$$

Finally, the Precision, Recall, and the F1-score metrics are evaluated as follows:

$$\begin{aligned} Precision_j &= \frac{TP_j}{TP_j + FP_j} \\ Recall_j &= \frac{TP_j}{TP_j + FN_j} \\ F_{1j} &= \frac{2 \cdot Precision_j \cdot Recall_j}{Precision_j + Recall_j} \end{aligned} \tag{27}$$

5. Results

The performance of the optimized LSTM AE in discriminating healthy and damaged structural conditions is assessed by applying the anomaly detection framework to four test scenarios. The same neural network architecture characterized by the BO-based hyperparameters is employed in all tests. The tests include both healthy and damaged conditions of the railway bridge (see Section 3), allowing to examine the sensitivity of the reconstruction-based indicators under progressively more severe structural alterations. Other than the three damage scenarios, a healthy (H) testing scenario is considered. This test is made to verify that the LSTM AE does not produce spurious detections when exposed to new realizations of the same healthy structural behavior, and to establish a reference level of reconstruction error against which damaged conditions can be meaningfully compared. For all the tests, the window length is 10 s, as described in Section 4.1. The selected time length represents a compromise between temporal resolution and the need to capture a sufficiently rich description of the bridge dynamic response during train passages. Specifically, the chosen duration allows for the inclusion of multiple cycles of the dominant bending and torsional modes, as well as the full transient response induced by the train–bridge interaction. Shorter windows may not fully capture the relevant dynamic behavior; longer windows would reduce temporal resolution without providing significant additional information.

Across the four scenarios considered in this paper, the results presented in the following include:

- The global reconstruction error E_i across time windows, together with the global anomaly threshold derived from healthy-condition training data, enabling the identification of windows whose overall dynamic response deviates from the healthy signature;
- The reconstruction error is compared between healthy and damaged scenarios by averaging, for each sensor, the RMSE over all the available time windows. This

representation provides insight into which sensors are most sensitive to the imposed damage and how the anomaly manifests spatially along the structure;

- The anomaly detection performance is quantified through sensor-wise RI, computed according to the statistical thresholds established in the training step. The resulting RI curves summarize, for each sensing point, the reliability with which structural deviations are detected across all windows, through a direct performance metric. Moreover, to enhance the interpretability of the results, the sensors are grouped according to their position along the upstream and downstream sides of the railway bridge, corresponding to odd-indexed and even-indexed sensors, respectively, as shown in Figure 2.

The results presented in this study have been obtained using simulated data derived from a calibrated finite element model of the railway bridge. This allowed for a controlled and physically consistent evaluation of the proposed methodology. In contrast, real-world factors such as measurement noise, track irregularities, and environmental variability are not considered, introducing a domain gap between simulated and real monitoring conditions. However, the proposed framework is based on learning the intrinsic behavior of the structure under healthy conditions through multichannel time-series data, which is expected to mitigate the impact of noise and variability typically observed in real monitoring scenarios. The methodology will be validated using experimental data acquired from the monitored bridge in future research. For the tests that consider the railway bridge in damaged conditions, the physical sequence of time windows adopted as input for the trained LSTM AE follows that presented in Figure 5.

5.1. Healthy Test

The internal consistency of the proposed anomaly detection framework is verified through the present test by applying the LSTM AE to vibration data corresponding to healthy structural conditions only. The time windows, belonging to the testing dataset described in Section 4, are randomly ordered. The capability of the network to correctly recognize healthy conditions without generating spurious alarms, and the definition of per-sensor and global thresholds from undamaged structural behavior, constitute the assessments of the present analysis.

The first result, shown in Figure 8, is provided by the global reconstruction error index E_i, reported as a function of the time window index. As shown in the upper plot, the global error remains consistently below the global threshold identified during training, resulting in a positive rate $PR_{glob} = 0$. Therefore, the LSTM AE does not produce false global anomaly detections under healthy conditions, despite the presence of natural variability induced by railway vehicle passages and operational excitation. Figure 8 reports the anomaly map obtained by applying the sensor-level criterion based on the three- sigma threshold $\mu_j + 3\sigma_j$. Only a very limited number of isolated threshold exceedances can be observed, affecting single sensors and single time windows. These exceedances do not show temporal persistence nor spatial correlation across neighboring sensors and therefore cannot be associated with structural damage. The global anomaly index remains consistently below the detection threshold, confirming the absence of false positives at the system level. The sensor-level map shows only isolated and non-persistent threshold exceedances, indicating stable behavior of the monitored structure and robustness of the proposed method under nominal operating conditions.

The per-sensor RI evaluated using the one-window criterion, as indicated in Equation (19), further supports this observation. For sensors located upstream of the railway bridge, the RI remains very close to unity across all sensor positions, with only marginal reductions for a limited number of sensors, as shown in Figure 9. Near-perfect

RI is achieved across all sensors, confirming the absence of false anomaly detections. The two-window criterion provides fully stable classification, eliminating minor fluctuations observed in the one-window approach and enhancing the robustness of the detection framework under nominal operating conditions. A similar trend is observed for the downstream sensor line. These results indicate that the network correctly classifies healthy windows at the sensor level, with negligible false positives.

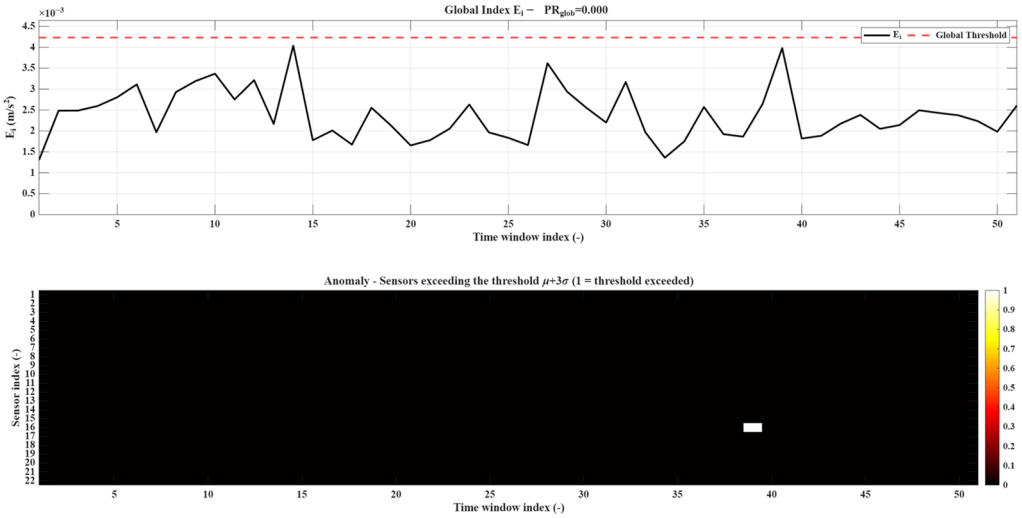

Figure 8. H Test: global reconstruction error E_i over time windows and corresponding sensor-level threshold exceedances under healthy conditions.

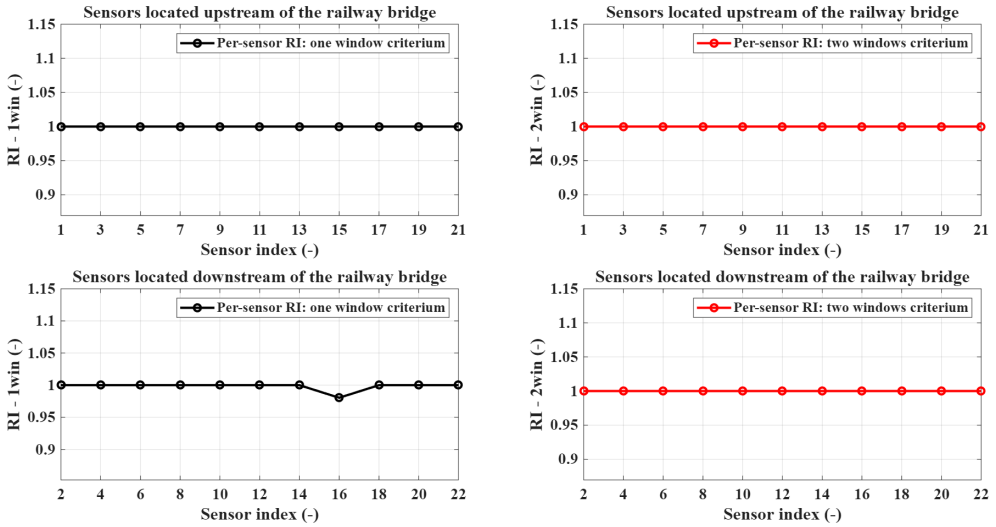

Figure 9. H Test: per-sensor RI under healthy conditions for upstream and downstream sensors using one-window and two-window anomaly detection criteria. On sensor 16, 2% of false positives are detected using the one-window criteria.

A more conservative behavior is obtained when the two-window consistency criterion is applied, as observable in Figure 9. In this case, the per-sensor RI, computed as described in Equation (22), reaches exactly 1 for both upstream and downstream sensors, meaning that no anomalies are declared when requiring two consecutive threshold exceedances. The two-window criterion effectively suppresses isolated false alarms. Furthermore, the sensitivity to persistent deviations is preserved.

Figure 10 reports, only for the present test, the sensor-level thresholds computed as indicated in Equation (15), directly derived from the healthy-condition dataset, described in Section 4.1. The thresholds exhibit a clear spatial variability along the bridge, reflect-

ing differences in local dynamic response, sensor position, and structural stiffness. This result highlights an important physical aspect of the proposed approach: the anomaly detection criteria are not based on uniform thresholds but are tailored to the local vibration characteristics of each sensor, thereby increasing reliability and interpretability.

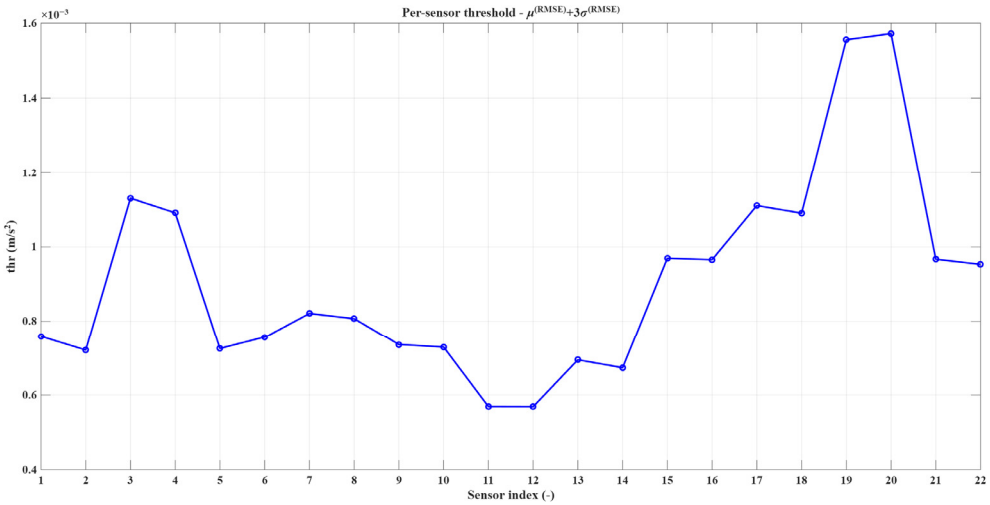

Figure 10. H Test: per-sensor reconstruction error thresholds computed under healthy structural conditions.

The proposed framework, based on the LSTM AE, is fully consistent under healthy conditions, without introducing artificial anomalies, and provides physically meaningful thresholds usable for subsequent damage detection analyses, as demonstrated by results obtained in the H Test.

5.2. GC Test

The GC Test is aimed at assessing the capability of the trained LSTM Autoencoder to detect a mild and spatially distributed structural modification, representative of a general corrosion scenario. In this test, the railway bridge is subjected to a uniform reduction in both Young's modulus and material density equal to 5%, affecting the entire structure.

From a physical standpoint, this type of damage is expected to induce limited variations in the global dynamic response of the bridge. Due to the distributed nature of stiffness and mass reduction and considering the overall inertial and structural characteristics of the system, no pronounced local effects are anticipated. Consequently, the bridge response remains globally coherent, and the induced damage cannot be considered critical from a structural safety perspective.

This behavior is clearly reflected in the global reconstruction error E_i, displayed in Figure 11. The global anomaly index frequently exceeds the detection threshold, indicating the presence of a distributed structural alteration. The sensor-level map shows widespread and persistent threshold exceedances across multiple sensors, reflecting the spatially uniform nature of the damage and confirming the capability of the proposed method to detect global changes in the structural response.

A more insightful interpretation emerges from the per-sensor reconstruction error analysis. Figure 12 shows the mean per-sensor RMSE values, separately reported for sensors located upstream and downstream of the railway bridge. An increase in reconstruction error is observed across all sensors under damaged conditions, with a consistent shift between healthy and damaged curves. The uniform distribution of the error increase confirms the spatially distributed nature of the damage and highlights the sensitivity of the proposed method to global structural alterations. Specifically, the healthy baseline has been

obtained using the dataset $W_{\text{healthy,TOT}}$, presented in Section 4.1, organized according to the sequence of time windows, shown in Figure 5, without a random permutation.

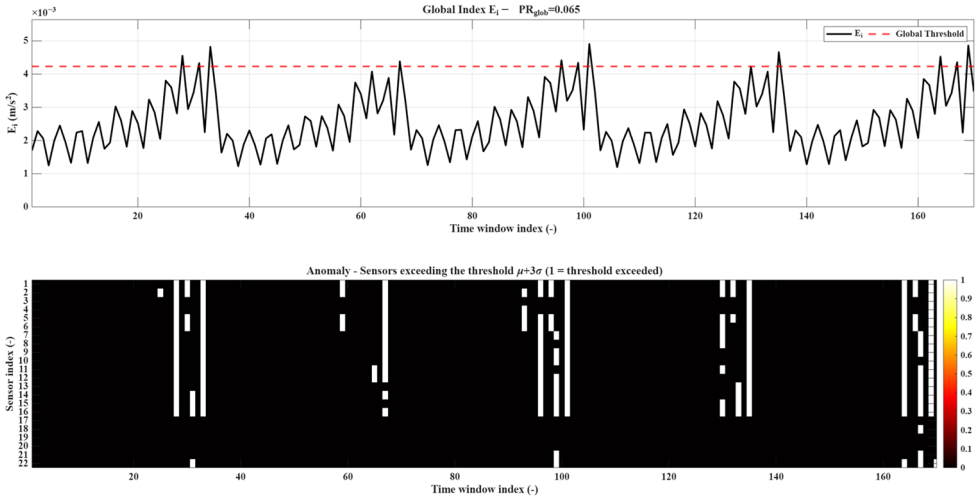

Figure 11. GC Test: global reconstruction error E_i over time windows and corresponding sensor-level threshold exceedances for the global corrosion (GC) scenario.

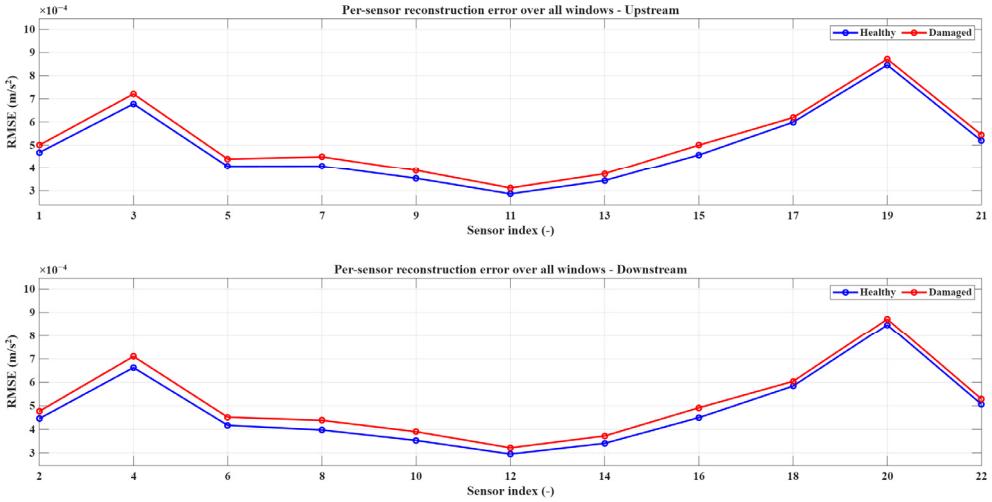

Figure 12. GC Test: per-sensor RMSE averaged over all time windows for upstream and downstream sensors in healthy and GC conditions.

A physically meaningful signature of general corrosion is represented by the nearly constant offset between the healthy and damaged mean RMSE curves. Localized damage scenarios present sharp spatial gradients. In contrast, general corrosion is manifested as a global alteration of structural dynamics. The global damage is captured by the LSTM AE, which produces systematically higher reconstruction errors without concentrating anomalies on specific sensor locations.

The per-sensor RI values further support this interpretation. For both upstream and downstream sensors, the RI associated with the one-window criterion, shown in Figure 13, remains close to unity, with only marginal reductions. This indicates that, although the anomaly is detectable at the reconstruction error level, its intensity is not sufficient to trigger a persistent or critical degradation in detection performance. Specifically, the one-window criterion shows moderate variability across sensors, reflecting the sensitivity of the detection to distributed structural changes. In contrast, the two-window criterion provides fully stable classification with perfect RI across all sensors, demonstrating the effectiveness of temporal consistency in improving robustness for global damage detection.

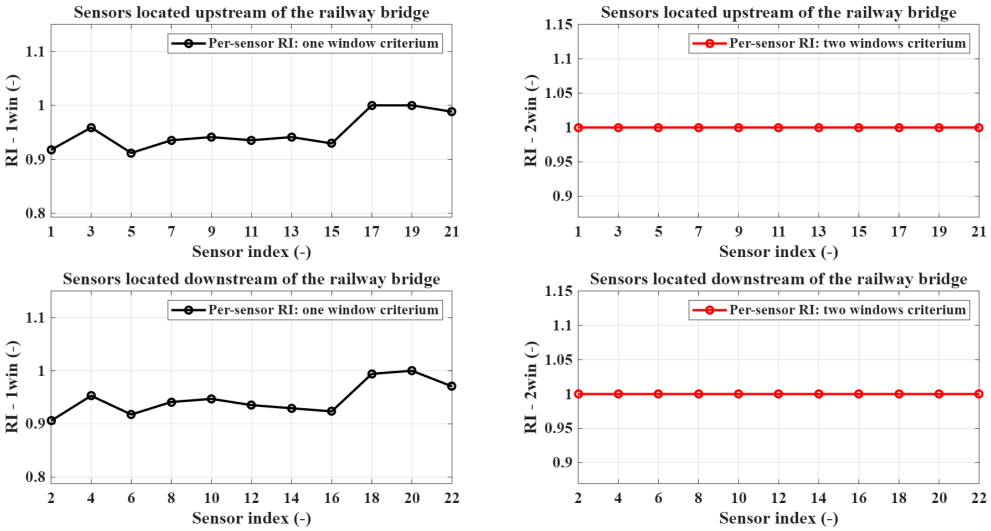

Figure 13. GC Test: per-sensor RI for the GC scenario using one-window and two-window anomaly detection criteria for upstream and downstream sensors.

By applying the stricter two-window consistency criterion, the RI reaches unity for all sensors, as shown in Figure 13, confirming the absence of a severe or alarming damage condition. The proposed LSTM AE can identify weak spatially distributed anomalies, preserving physical consistency, as demonstrated by the results of the GC Test.

5.3. LD Test10

The response of the proposed anomaly detection framework, based on the LSTM AE, in the presence of a localized damage of limited severity, is investigated in the LD Test10. The damage produces only mild alterations in the global dynamic behavior of the railway bridge, due to its small extent. Therefore, the damage is weakly perceived at the structural level. The global reconstruction error E_i confirms this aspect, remaining largely below the global threshold for most of the analyzed time windows, as shown in Figure 14. Only a small number of exceedances are observed, resulting in a low global positive rate, indicating that the introduced damage does not significantly compromise the overall structural response. The global anomaly index occasionally approaches or slightly exceeds the detection threshold, indicating mild and intermittent deviations from nominal behavior. The sensor-level map shows sparse and localized threshold exceedances concentrated around specific sensors, reflecting the limited spatial extent and low severity of the damage.

The RMSE profiles are shown in Figure 15. An increase in RMSE can be observed in correspondence with the sensors located near the damage position, with the most significant variations detected around the affected region. The other sensors, which are located away from the area near sensor 8, exhibit RMSE values of the reconstruction error comparable to healthy conditions, consistent with the localization of the damage in the structure. Furthermore, adjacent sensors are affected by the increase in reconstruction error, which is not strictly limited to a single sensor, demonstrating a consistent spatial propagation of structural vibrations. This behavior reflects the dynamic coupling of the bridge structure, where localized stiffness variations influence the response over a neighborhood of sensors. The observed pattern confirms the capability of the proposed approach to capture localized damage, preserving consistency with the underlying physical behavior of the system. This result confirms that, despite the limited severity of the damage, the proposed approach can isolate localized alterations without inducing spurious global detections.

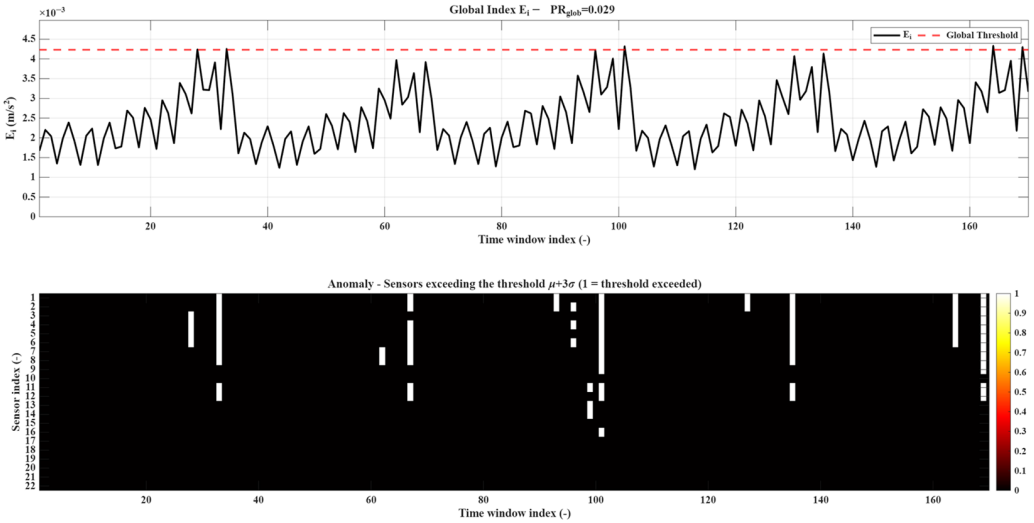

Figure 14. LD Test10: global reconstruction error E_i over time windows and corresponding sensor-level threshold exceedances for the LD10 scenario.

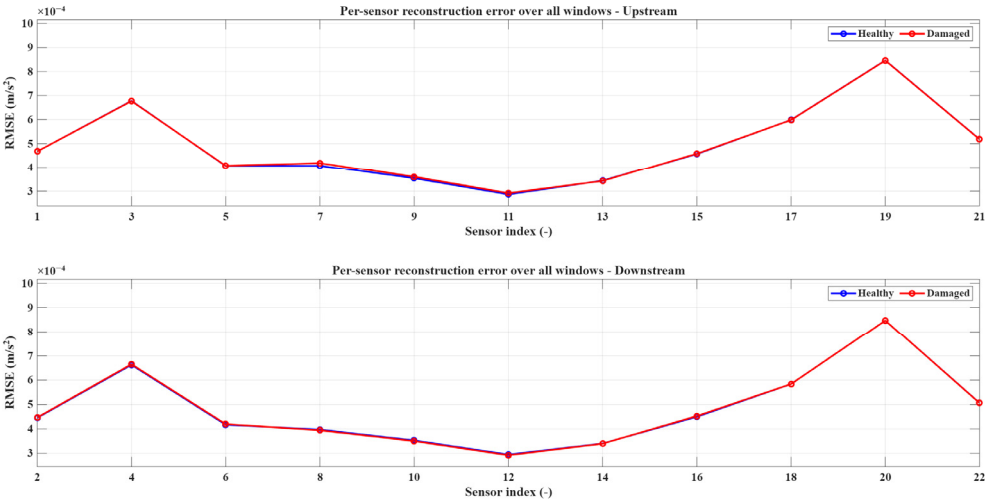

Figure 15. LD Test10: per-sensor RMSE averaged over all time windows for upstream and downstream sensors in healthy and LD10 conditions.

The per-sensor RI analysis indicates high RI levels for both the one-window and two-window consistency criteria, shown in Figure 16. Minor RI reductions are observed in correspondence with the affected damaged region, while all other locations maintain values close to unity. High RI is maintained across all sensors, indicating that the mild localized damage does not significantly affect the overall detection performance. The one-window criterion shows only minor variability, while the two-window criterion ensures fully stable classification, confirming the robustness of the proposed method under low-severity damage conditions.

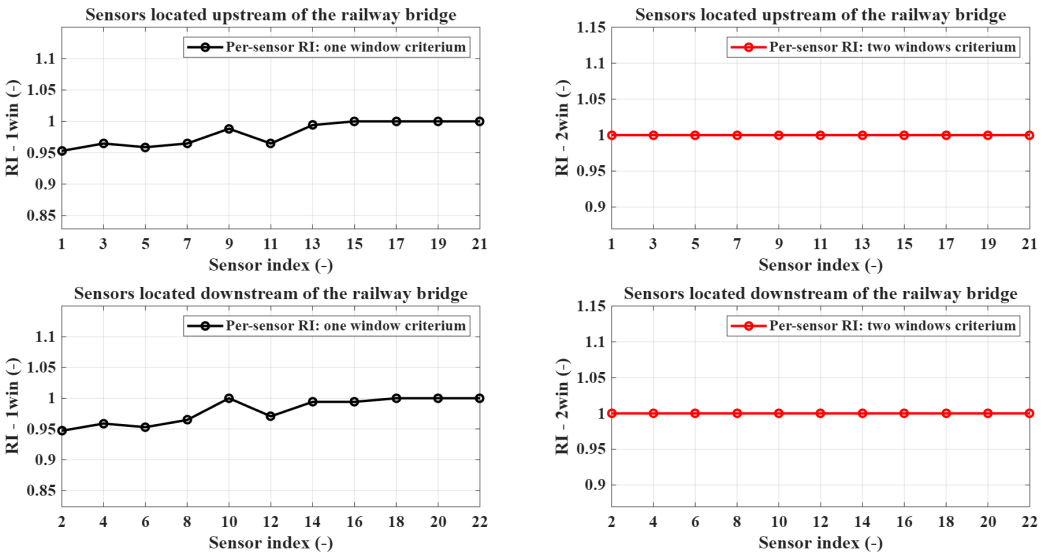

Figure 16. LD Test10: Per-sensor RI for the LD10 scenario using one-window and two-window anomaly detection criteria for upstream and downstream sensors.

5.4. LD Test70

The LD Test70 scenario represents a severe localized damage condition, characterized by a pronounced stiffness and density reduction concentrated in a limited portion of the structure. Although the global dynamic response of the bridge still preserves a periodic pattern driven by train passages, the presence of damage induces clear and spatially coherent anomalies in the reconstruction error metrics.

The global reconstruction index, E_i, shown in Figure 17, reports a marked increase in the number of threshold exceedances compared to mild damage cases. The global anomaly index frequently exceeds the detection threshold, indicating pronounced deviations from nominal structural behavior. The sensor-level map reveals widespread and persistent threshold exceedances, particularly concentrated around the damaged region and extending to neighboring sensors, reflecting the increased severity of the damage and its broader spatial influence on the structural response.

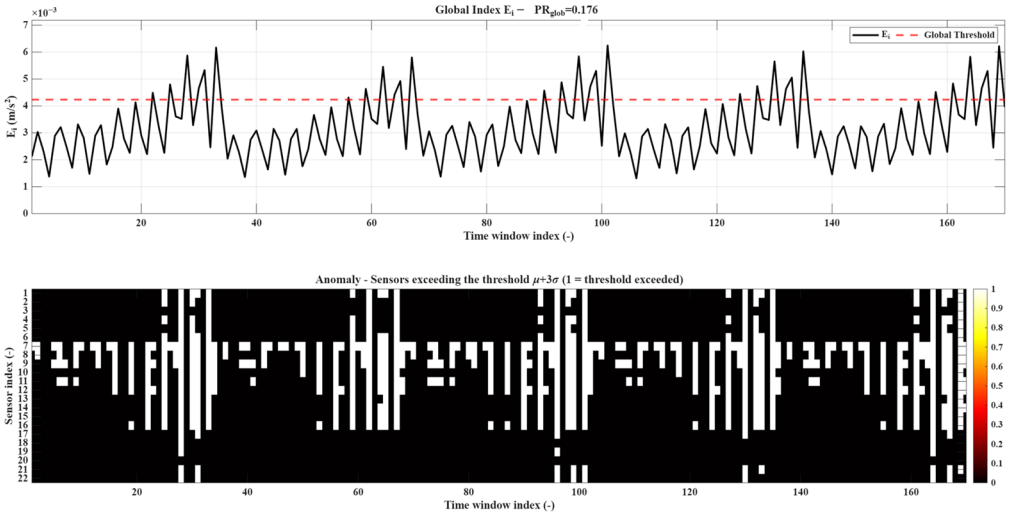

Figure 17. LD Test70: global reconstruction error E_i over time windows and corresponding sensor-level threshold exceedances for the LD70 scenario.

The presence of severe local damage can be further observed in the per-sensor mean RMSE comparison presented in Figure 18. A significant offset between healthy and dam-

aged conditions is highlighted in the region containing sensor 8, near the applied local damage. The anomaly is confined to the previously mentioned region, and not to a single sensor. This is consistent with the increased stiffness reduction, which induces a more pronounced modification of the global dynamic response of the railway bridge. The offset is sharply concentrated, consistent with the physical characteristics of a localized damage mechanism. As in the LD10 case, the highest reconstruction errors are observed in proximity to the damaged region, while adjacent sensors also exhibit elevated values due to the spatial propagation of vibrations. Furthermore, a non-uniform offset is observable, unlike global corrosion scenarios.

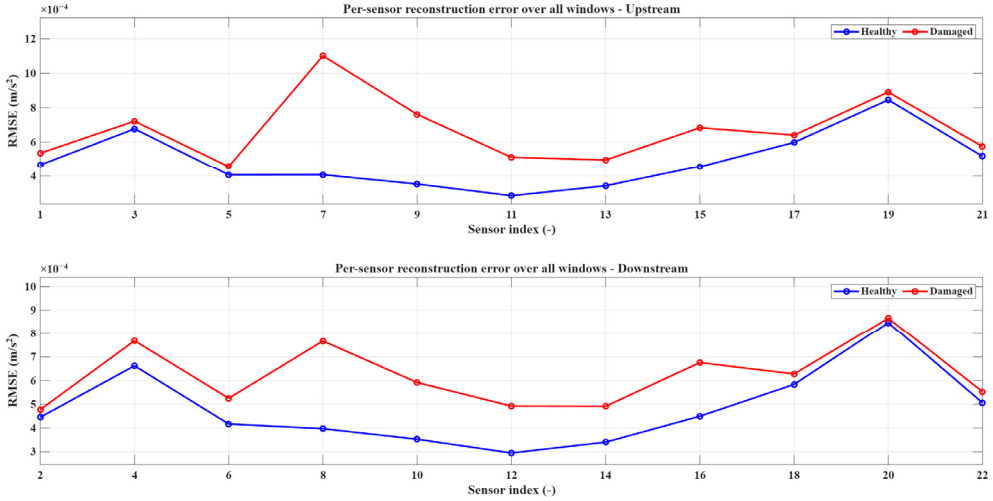

Figure 18. LD Test70: per-sensor RMSE averaged over all time windows for upstream and downstream sensors in healthy and LD70 conditions.

The impact of damage severity is reflected in the per-sensor RI metrics. Sensors close to the damage region exhibit a substantial drop in RI, with sensors 7 and 8 reaching the lowest values among the network, indicating a consistent classification of damaged windows as anomalous, as expected in the presence of strong local deviations, as shown in Figure 19, observing the one-window criterion.

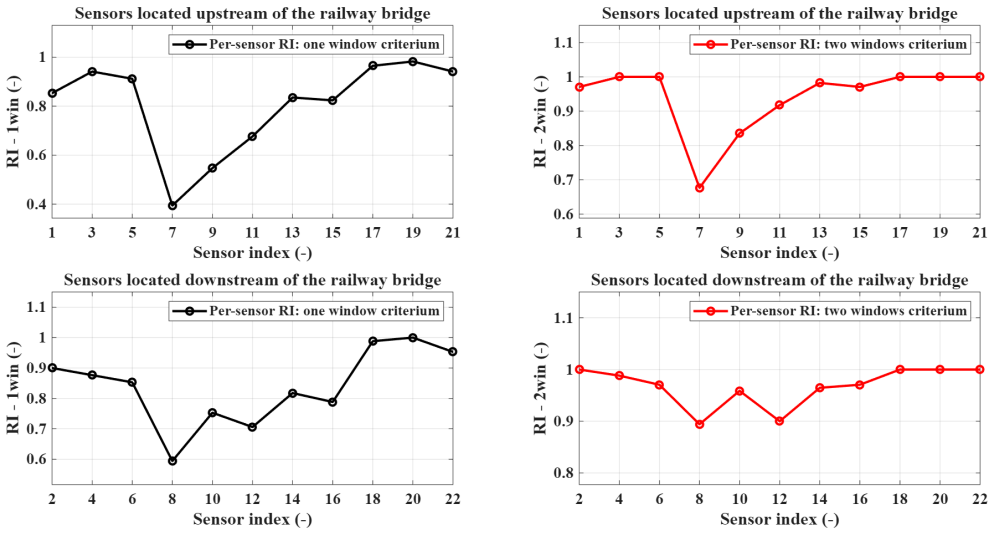

Figure 19. LD Test70: per-sensor RI for the LD70 scenario using one-window and two-window anomaly detection criteria for upstream and downstream sensors.

Furthermore, a noticeable reduction in RI is highlighted at sensors 7 and 8 when applying the two-window consistency criterion, coupled with an overall RI improvement for most sensors, as shown in Figure 19, confirming both the reliability of the detection strategy and the persistence of damage-related effects.

The LSTM AE is suitable for capturing both the severity and the spatial localization of damage, as demonstrated by the LD Test70. The clear RMSE offset in the region containing sensors 7 and 8, and the coherent degradation of RI collectively indicate a physically meaningful and reliable detection of a severe localized structural anomaly. Furthermore, this behavior confirms that more pronounced damage scenarios lead to a more distributed and complex alteration of the structural dynamics, which is effectively captured by the proposed methodology.

Finally, the Precision, Recall, and F1-score, computed as presented in Section 4.3, are summarized in Figure 20 for all the scenarios considered. In the present application, these metrics indicate the detectability of the damage across the sensor network, rather than their classical adoption for classification problems. The results are consistent with the trends observed in the RI-based analysis. In the healthy condition, all sensors exhibit null recall and F1-score (as there are no positives to be detected), confirming the absence of false damage detections and the reliability of the method in nominal conditions. Considering the scenarios containing information on damages, the behavior of the F1-score reflects the spatial detectability of the damage of the proposed technique. In the GC-Test, the F1-score remains close to one for almost all sensors, consistent with the uniformly distributed imposed defect represented by a general corrosion of the railway bridge. The corresponding RI values, which deviate from unity across the sensor network, are in accordance with the F1-score.

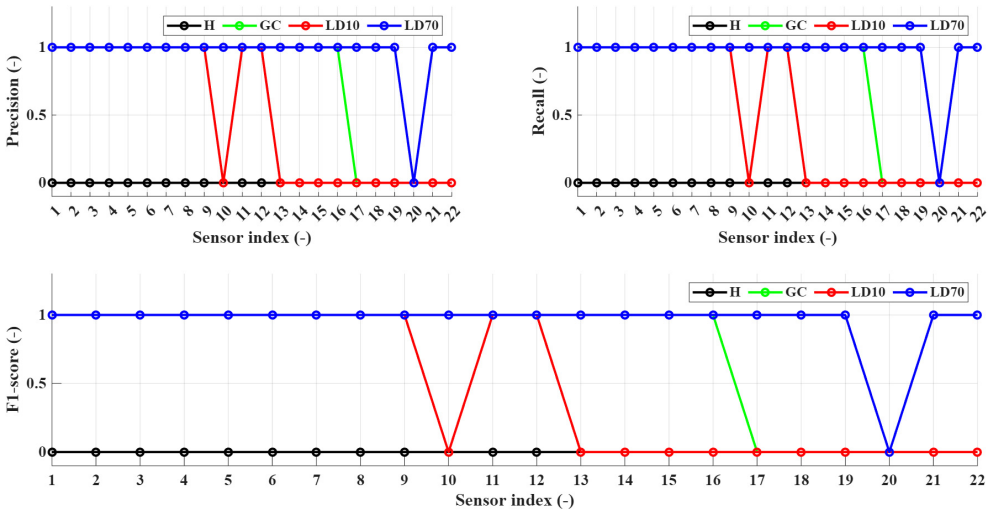

Figure 20. Sensor-level Precision, Recall, and F1-score for all considered scenarios, highlighting the spatial detectability of damage across the sensor network.

In contrast, for the LD10 scenario, the F1-score exhibits a more heterogeneous pattern, with several sensors displaying null values, consistent with the corresponding RI, where only a subset of sensors presents a significant deviation from the healthy condition. Therefore, the presence of local damage is confirmed, coupled with a weak excitation of the railway bridge, making it only partially observable. Finally, in the LD70 scenario, the F1-score confirms a nearly global detectability of the damage, with almost all sensors achieving perfect detection (F1 = 1), except for sensor 20. The related RI results are coherent with the F1-score, showing a widespread deviation from the healthy condition, indicating a clear observability of the damage across the structure. Furthermore, the RI analysis enables

accurate localization of the damage, highlighting the region around sensor 8, with sensors 7 and 8 exhibiting the most significant response, as previously discussed.

The obtained results demonstrate the potential of the proposed approach under controlled simulation conditions. However, the absence of measurement noise and track irregularities represents a limitation that may affect the direct transferability of the method to real monitoring scenarios. The inclusion of the previously mentioned real-world features affecting the railway bridge and the sensing instrumentation will be considered in future studies related to the present work.

6. Conclusions

An unsupervised monitoring methodology for railway bridges based on an LSTM AE trained exclusively on healthy-condition acceleration data has been presented in this work. The methodology exploits multichannel acceleration signals, obtained from a detailed numerical model simulating the coupled train–track dynamics, and a reconstruction error-based indicator to identify deviations from nominal structural behavior. The reliability of the proposed monitoring approach in distinguishing between healthy conditions, mild global degradation, and localized damage of increasing severity has been demonstrated by the results obtained on the different test scenarios. The results obtained from the numerical simulations demonstrate that the proposed approach captures changes in the structural dynamic response through variations in reconstruction error. Particularly, the method demonstrates sensitivity to both distributed and localized damage scenarios. Furthermore, the adopted dual-threshold strategy enables distinguishing between global anomalies and sensor-level effects, confirming the suitability of the proposed approach for continuous monitoring applications, where reliability and avoidance of false alarms are essential. Overall, the proposed methodology, based on the LSTM AE, is suitable for structural health monitoring of railway bridges, representing a physically consistent, data-driven, and computationally efficient tool capable of operating under realistic operational variability, without requiring labeled damaged data or numerical model updating.

Future developments will focus on investigating more articulated and computationally efficient autoencoder architectures, aiming to enhance sensitivity to early-stage damage and reduce training and inference costs. Additionally, damage classification criteria explicitly linked to the structural characteristics of the bridge will be defined, enabling both anomaly detection and a more direct interpretation of damage type and severity within a bridge-specific diagnostic framework. Furthermore, more advanced statistical approaches related to thresholding will be explored to further improve detection robustness, including confidence interval estimation and probabilistic modeling of reconstruction errors. Finally, it is important to stress that train–track–bridge dynamic simulations will be carried out, accounting also for track geometrical profile and its evolution in time, to further investigate the robustness of the proposed approach, including validations based on experimental data collected on the monitored bridge.

Author Contributions: Conceptualization, V.G., C.T. and L.B.; methodology, V.G. and C.T.; software, C.T.; investigation, C.T.; data curation, V.G. and P.A.R.B.; writing—original draft preparation, V.G. and C.T.; writing—review and editing, L.B.; supervision, C.S. and S.S.; project administration, C.S., S.S. and M.T. All authors have read and agreed to the published version of the manuscript.

Funding: Financed by the European Union—Next Generation EU, Centro Nazionale Mobilità Sostenibile (MOST), Scalability Project "BE-REMOTE" (Spoke 4), CN00000023, M4C2 I1.4, CUP D43C22001180001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are confidential and are therefore not publicly available.

Acknowledgments: The authors would like to acknowledge the company Ferrovie Nord S.p.A. for the support during preliminary tests and the installation of the monitoring system.

Conflicts of Interest: Author Pablo Alex Ramírez Balbiano is employed by the company Displaid SRL Società Benefit. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

Appendix A.1. Structural Properties of the FE Model

This Section focuses on the structural properties of the bridge after the calibration process. The steel density was increased by 20% to account for the mass contributions related to elements not considered in beams' net cross-sections (e.g., plates and bolted connections) as well as non-structural components. Table A1 shows the Young's Modulus, Poisson's Ratio and density of both the steel and concrete materials used to model the bridge, as well as the obtained Rayleigh damping coefficients.

Table A1. Main structural properties used in the bridge model. The density of the steel (*) was increased by 20% during the calibration phase.

Bridge Structure			
Material	Young's Modulus [MPa]	Poisson's Ratio [-]	Density [kg/m^3]
Steel	206×10^3	0.3	9420 *
Concrete	34.07×10^3	0.2	2549
Proportional damping parameters			
α_0 (mass-proportional damping)		$0.517\ s^{-1}$	
α_1 (stiffness-proportional damping)		0.001 s	
Track structure			
Rail section [m^2]		0.767×10^{-2}	
Rail mass per unit length [kg/m]		60	
Vertical rail pad stiffness [MN]		354	

Table A2 shows the stiffness of the elements modeled as springs, i.e., the Gerber saddles and the bridge piers. Table A3 shows the damping coefficients of these elements. Figure A1 shows the placement of the different Gerber saddles on the central span of the bridge.

Table A2. Stiffness of the Gerber saddle and pier springs.

Spring ID	**k_x [N/m]**	**k_y [N/m]**	**k_z [N/m]**	**Description**
S1–S3	2×10^8	2×10^{12}	2×10^{12}	Gerber saddles 1–3, in red in Figure A1
S4	2×10^{12}	2×10^{12}	2×10^{12}	Gerber saddle 4, in yellow in Figure A1
S5	-	2×10^{12}	1×10^9	Pier supports downstream
S6	-	-	1×10^9	Pier supports upstream

Table A3. Damping coefficients of the Gerber saddle and pier springs.

Spring ID	c_x [N·s/m]	c_y [N·s/m]	c_z [N·s/m]	Description
S1–S3	2×10^5	2×10^9	2×10^9	Gerber saddles 1–3, in red in Figure A1
S4	2×10^9	2×10^9	2×10^9	Gerber saddle 4, in yellow in Figure A1
S5	-	2×10^9	1×10^6	Pier supports downstream
S6	-	-	1×10^6	Pier supports upstream

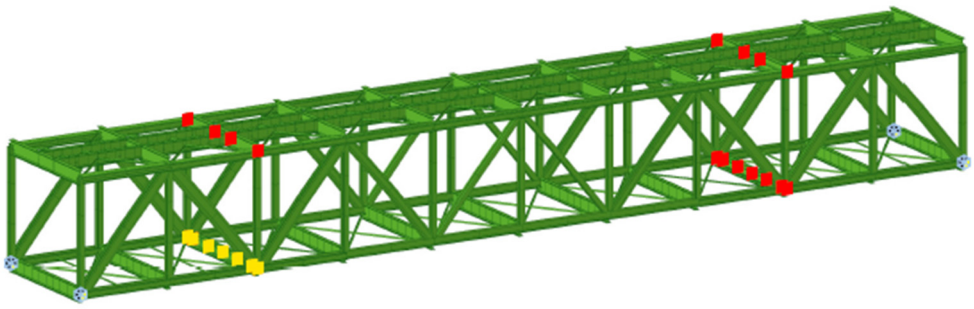

Figure A1. Placement of the Gerber saddles. S1–S3 are drawn in red, while S4 is in yellow.

Appendix A.2. Numerical Mode Shapes

This section of Appendix A presents the numerical mode shapes after the calibration of the FE model, pictured in Figures A2–A6.

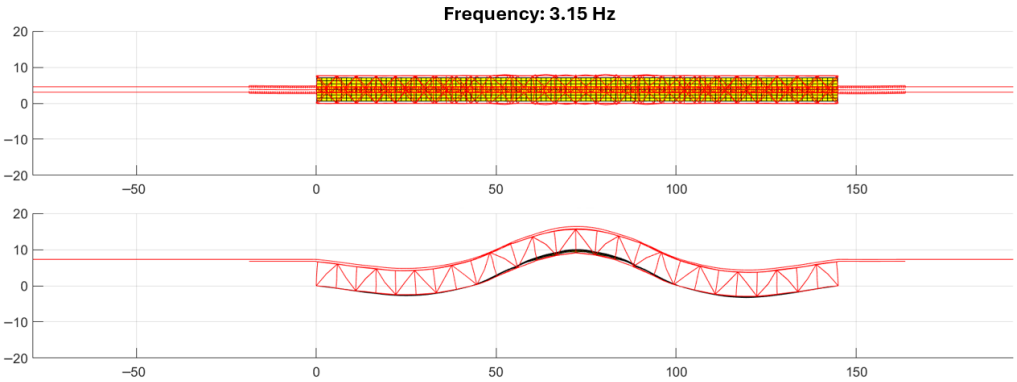

Figure A2. First vertical mode shape after calibration, at 3.15 Hz. The (**upper**) figure is the top view of the Y axis, and the (**bottom**) figure is the lateral view of the Z axis.

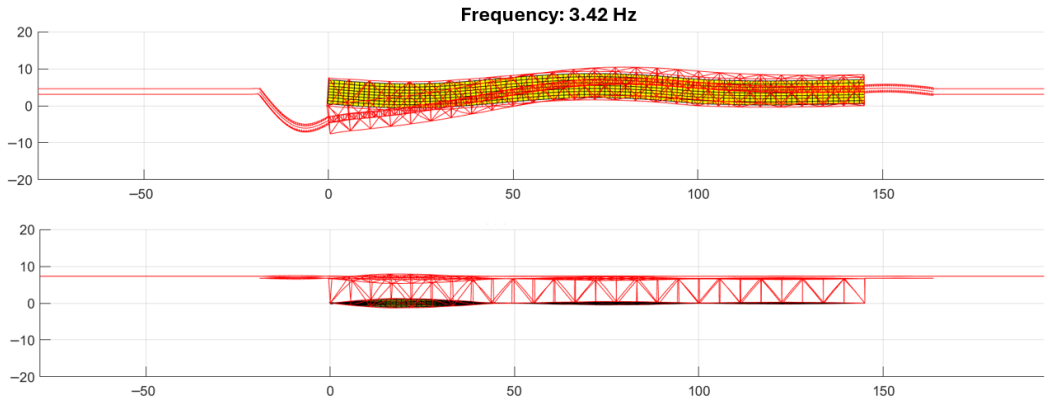

Figure A3. First lateral mode shape after calibration, at 3.42 Hz. The (**upper**) figure is the top view of the Y axis, and the (**bottom**) figure is the lateral view of the Z axis.

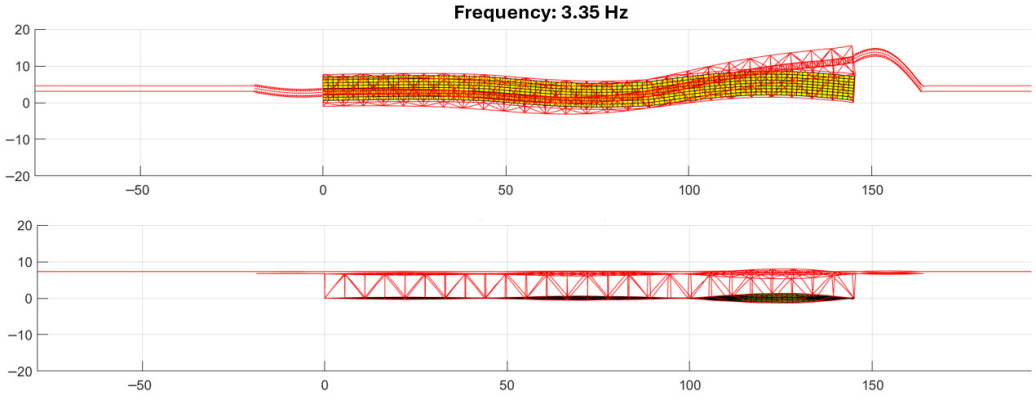

Figure A4. Second lateral mode shape after calibration, at 3.35 Hz. The (**upper**) figure is the top view of the Y axis, and the (**bottom**) figure is the lateral view of the Z axis.

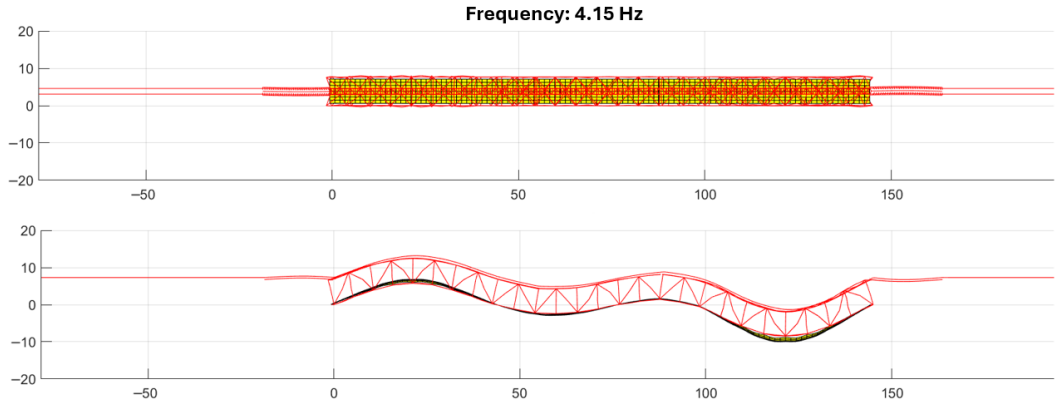

Figure A5. Second vertical mode shape after calibration, at 4.15 Hz. The (**upper**) figure is the top view of the Y axis, and the (**bottom**) figure is the lateral view of the Z axis.

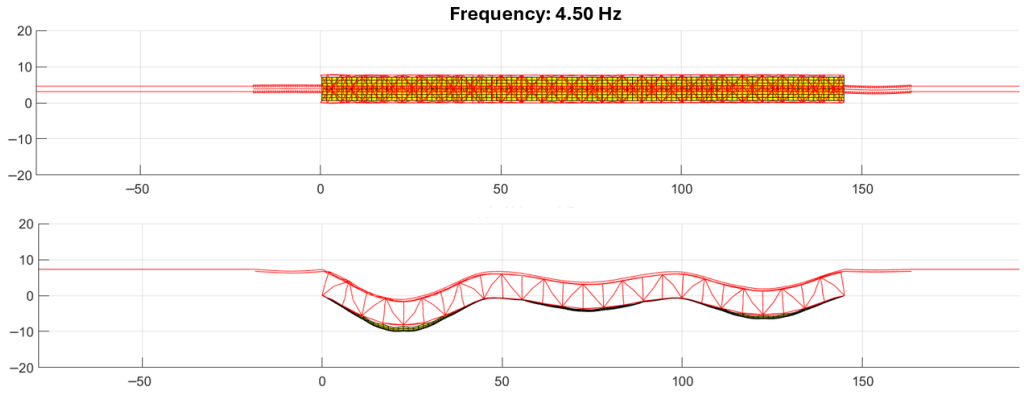

Figure A6. Third vertical mode shape after calibration, at 4.5 Hz. The (**upper**) figure is the top view of the Y axis, and the (**bottom**) figure is the lateral view of the Z axis.

Appendix B

Appendix B.1. Architecture of the LSTM AE

This section describes the theory behind the architecture of the LSTM AE used in the present work [49,50].

The input signal for the LSTME AE is denoted as follows:

$$\boldsymbol{X}_i = \{\mathbf{x}_t\}_{t=1}^{L}.\text{with } \mathbf{x}_t \in \mathbb{R}^{1\times 22}; i = 1, \ldots, n_W \quad \text{(A1)}$$

where $\mathbf{x}_t$ represents the 22-dimensional vector of vertical accelerations measured simultaneously at time t, L is the window length expressed in samples, and n_W is the number of time windows. The LSTM AE aims at learning a nonlinear mapping:

$$\boldsymbol{F}_\theta : \boldsymbol{X} \mapsto \hat{\boldsymbol{X}} \tag{A2}$$

where $\boldsymbol{\theta}$ denotes the complete set of trainable parameters of the network, including all LSTM weight matrices, recurrent kernels and bias terms, such that the reconstructed sequence $\hat{\boldsymbol{X}} = \{\hat{\mathbf{x}}_t\}_{t=1}^{L}$ approximates the healthy structural response with high fidelity.

The neural network is composed of three main components: the encoder, the bottleneck, and the decoder, represented by three Long Short-Term Memory (LSTM) layers, respectively. The encoder consists of a recurrent layer with N_{enc} units, represented by an even number.

The temporal evolution is governed by the input gate $\mathbf{i}_t$, the forget gate $\mathbf{f}_t$, the output gate $\mathbf{o}_t$, the candidate cell state $\tilde{\mathbf{c}}_t$, the internal memory state $\mathbf{c}_t$, and the hidden state $\mathbf{h}_t$ in a LSTM cell at each time step t.

The previously mentioned characteristic variables of the first LSTM cell, represented by the encoder layer, evolve according to the classical LSTM relations [51]:

$$\begin{gathered}\mathbf{i}_t^{(1)} = \sigma\left(\mathbf{W}_i^{(1)}\mathbf{x}_t + \mathbf{U}_i^{(1)}\mathbf{h}_{t-1}^{(1)} + \mathbf{b}_i^{(1)}\right), \mathbf{f}_t^{(1)} = \sigma\left(\mathbf{W}_f^{(1)}\mathbf{x}_t + \mathbf{U}_f^{(1)}\mathbf{h}_{t-1}^{(1)} + \mathbf{b}_f^{(1)}\right) \\ \mathbf{o}_t^{(1)} = \sigma\left(\mathbf{W}_o^{(1)}\mathbf{x}_t + \mathbf{U}_o^{(1)}\mathbf{h}_{t-1}^{(1)} + \mathbf{b}_o^{(1)}\right), \tilde{\mathbf{c}}_t^{(1)} = \tanh\left(\mathbf{W}_c^{(1)}\mathbf{x}_t + \mathbf{U}_c^{(1)}\mathbf{h}_{t-1}^{(1)} + \mathbf{b}_c^{(1)}\right) \\ \mathbf{c}_t^{(1)} = \mathbf{f}_t^{(1)} \odot \mathbf{c}_{t-1}^{(1)} + \mathbf{i}_t^{(1)} \odot \tilde{\mathbf{c}}_t^{(1)}, \mathbf{h}_t^{(1)} = \mathbf{o}_t^{(1)} \odot \tanh\left(\mathbf{c}_t^{(1)}\right)\end{gathered} \tag{A3}$$

where $\mathbf{W}_{\cdot}^{(1)}$ and $\mathbf{U}_{\cdot}^{(1)}$ are trainable weight matrices mapping, respectively, the current input $\mathbf{x}_t$ and the previous hidden state $\mathbf{h}_{t-1}^{(1)}$, $\mathbf{b}_{\cdot}^{(1)}$ are the associated bias vectors, $\sigma(\cdot)$ denotes the logistic sigmoid, $\tanh(\cdot)$ the hyperbolic tangent, and $\odot$ the element-wise product.

In all LSTM layers, the logistic sigmoid $\sigma(\cdot)$ and the hyperbolic tangent $\tanh(\cdot)$ are employed as activation functions for the gating mechanisms and the candidate state update, respectively. The sigmoid maps gate activations into the interval $[0, 1]$, regulating the flow of information through the input, forget and output gates. The cell content and hidden state are constrained within $[-1, 1]$ by the hyperbolic tangent, providing a stable, bounded representation of the temporal dynamics.

The latent-space encoding generated by the first LSTM layer is subsequently processed through a dropout operator with rate p_{enc}, which randomly suppresses a fraction of the hidden-state components to mitigate overfitting and improve the generalization ability of the LSTM AE. The dropout-altered hidden state is denoted as follows:

$$\tilde{\mathbf{h}}_t^{(1)} = \mathbf{m}_t \odot \mathbf{h}_t^{(1)}, \mathbf{m}_t \sim \text{Bernoulli}(1 - p_{enc}) \tag{A4}$$

where $\mathbf{m}_t$ is a binary mask. The modified sequence $\tilde{\mathbf{h}}_t^{(1)}$ is the input for the second LSTM layer, which constitutes the bottleneck of the architecture.

The latent sequence is defined as follows by the compressed representation refined through the bottleneck layer having $N_{bneck} = 0.5 \cdot N_{enc}$ units:

$$Z = \{\mathbf{z}_t\}_{t=1}^{T}, \mathbf{z}_t = \mathbf{h}_t^{(2)} \in \mathbb{R}^{N_{bneck}}. \tag{A5}$$

The dynamics of the second layer replicate those of the first, with weight matrices $\mathbf{W}_{\cdot}^{(2)}$, $\mathbf{U}_{\cdot}^{(2)}$, and bias vectors $\mathbf{b}_{\cdot}^{(2)}$, by mapping from the dropout-altered encoder output to the corresponding cell and hidden states. The latent representation $\mathbf{z}_t$ constitutes a compact

description of the healthy structural behavior, in which only the most informative temporal patterns are retained.

The evolution of each variable is made according to the gating mechanisms indicated as follows coupled with the candidate cell state $\tilde{\mathbf{c}}_t^{(2)}$:

$$\begin{aligned} \mathbf{i}_t^{(2)} &= \sigma\left(\mathbf{W}_i^{(2)}\tilde{\mathbf{h}}_t^{(1)} + \mathbf{U}_i^{(2)}\mathbf{h}_{t-1}^{(2)} + \mathbf{b}_i^{(2)}\right), \mathbf{f}_t^{(2)} = \sigma\left(\mathbf{W}_f^{(2)}\tilde{\mathbf{h}}_t^{(1)} + \mathbf{U}_f^{(2)}\mathbf{h}_{t-1}^{(2)} + \mathbf{b}_f^{(2)}\right) \\ \mathbf{o}_t^{(2)} &= \sigma\left(\mathbf{W}_o^{(2)}\tilde{\mathbf{h}}_t^{(1)} + \mathbf{U}_o^{(2)}\mathbf{h}_{t-1}^{(2)} + \mathbf{b}_o^{(2)}\right), \tilde{\mathbf{c}}_t^{(2)} = tanh\left(\mathbf{W}_c^{(2)}\tilde{\mathbf{h}}_t^{(1)} + \mathbf{U}_c^{(2)}\mathbf{h}_{t-1}^{(2)} + \mathbf{b}_c^{(2)}\right) \end{aligned} \tag{A6}$$

The internal memory and the latent-state output are obtained as follows:

$$\mathbf{c}_t^{(2)} = \mathbf{f}_t^{(2)} \odot \mathbf{c}_{t-1}^{(2)} + \mathbf{i}_t^{(2)} \odot \tilde{\mathbf{c}}_t^{(2)}, \mathbf{h}_t^{(2)} = \mathbf{o}_t^{(2)} \odot \tanh\left(\mathbf{c}_t^{(2)}\right) \tag{A7}$$

A second dropout operation is applied to the latent representation provided by the bottleneck layer:

$$\tilde{\mathbf{z}}_t = \mathbf{n}_t \odot \mathbf{z}_t, \mathbf{n}_t \sim \text{Bernoulli}(1 - p_{bneck}) \tag{A8}$$

where p_{bneck} is the rate of the dropout operator applied below the bottleneck layer, which further regularizes the decoding stage, made by the third layer called decoder.

The original 22-dimensional sequence is reconstructed by the decoder, starting from the latent variables provided by the dropout-altered bottleneck layer. The decoder is an LSTM layer with $N_{dec} = N_{enc}$ units, which transforms each latent vector $\tilde{\mathbf{z}}_t$ into its decoded representation $\mathbf{h}_t^{(3)}$.

The gating equations of the decoder mirror those of the encoder as follows, by considering the parameters $\mathbf{W}_\cdot^{(3)}$, $\mathbf{U}_\cdot^{(3)}$, and $\mathbf{b}_\cdot^{(3)}$:

$$\begin{aligned} \mathbf{i}_t^{(3)} &= \sigma\left(\mathbf{W}_i^{(3)}\tilde{\mathbf{z}}_t + \mathbf{U}_i^{(3)}\mathbf{h}_{t-1}^{(3)} + \mathbf{b}_i^{(3)}\right), \mathbf{f}_t^{(3)} = \sigma\left(\mathbf{W}_f^{(3)}\tilde{\mathbf{z}}_t + \mathbf{U}_f^{(3)}\mathbf{h}_{t-1}^{(3)} + \mathbf{b}_f^{(3)}\right) \\ \mathbf{o}_t^{(3)} &= \sigma\left(\mathbf{W}_o^{(3)}\tilde{\mathbf{z}}_t + \mathbf{U}_o^{(3)}\mathbf{h}_{t-1}^{(3)} + \mathbf{b}_o^{(3)}\right), \tilde{\mathbf{c}}_t^{(3)} = \tanh\left(\mathbf{W}_c^{(3)}\tilde{\mathbf{z}}_t + \mathbf{U}_c^{(3)}\mathbf{h}_{t-1}^{(3)} + \mathbf{b}_c^{(3)}\right) \\ \mathbf{c}_t^{(3)} &= \mathbf{f}_t^{(3)} \odot \mathbf{c}_{t-1}^{(3)} + \mathbf{i}_t^{(3)} \odot \tilde{\mathbf{c}}_t^{(3)}, \mathbf{h}_t^{(3)} = \mathbf{o}_t^{(3)} \odot \tanh\left(\mathbf{c}_t^{(3)}\right) \end{aligned} \tag{A9}$$

where $\mathbf{h}_t^{(3)} \in \mathbb{R}^{N_{dec}}$ represents the decoded temporal feature vector at time t.

Finally, the reconstructed acceleration vector at time t is obtained through a linear mapping, for ensuring that the model accurately reproduces the dynamics of the healthy bridge response. Once trained exclusively on healthy data, the LSTM produces low reconstruction errors when the input reflects nominal structural behavior. Global or localized deviations in the structural status manifest themselves as increased discrepancies between $\mathbf{x}_t$ and $\hat{\mathbf{x}}_t$. The discrepancies, reflected in the reconstruction error, provide a physically meaningful basis for anomaly detection. The linear transformation is as follows:

$$\hat{\mathbf{x}}_t = \mathbf{W}_y\mathbf{h}_t^{(3)} + \mathbf{b}_y \tag{A10}$$

where $\mathbf{W}_y \in \mathbb{R}^{22\times N_{dec}}$ and $\mathbf{b}_y \in \mathbb{R}^{22}$ are learnable parameters mapping the decoder hidden state to the sensor domain.

Given the reconstructed sequence $\hat{\mathbf{X}} = \{\hat{\mathbf{x}}_t\}_{t=1}^{L}$, the LSTM AE is trained to approximate the healthy structural dynamics by minimizing the mean-squared reconstruction error (MSE) as follows:

$$\mathscr{L}(\boldsymbol{\theta}) = \frac{1}{L}\sum_{t=1}^{L} \| \hat{\mathbf{x}}_t - \mathbf{x}_t \|_2^2 \tag{A11}$$

where $\boldsymbol{\theta}$ denotes the complete set of trainable parameters of the network, including all LSTM weight matrices, recurrent kernels, bias terms, and the linear-output parameters.

Appendix B.2. Bayesian Optimization

This Section describes the procedure followed for the Bayesian Optimization of the LSTM AE parameters.

For a given hyperparameter vector $x \in \mathcal{X}$, where $\mathcal{X}$ denotes the hyperparameter search space, the LSTM AE is trained on the standardized windows $W_{\text{healthy,std-perm}}$, and the associated performance metric is defined as the average Root Mean Square Error (RMSE) computed over all reconstructed windows. By denoting with $\hat{W}_i$ the reconstruction of window W_i, the objective function is expressed as:

$$f(x) = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} \sqrt{\frac{1}{L} \sum_{k=1}^{L} \| W_i(k,22) - \hat{W}_i(k,22) \|_2^2} \tag{A12}$$

with $N_{\text{train}} = 170$ and $L = 5000$. Since the evaluation of $f(x)$ requires a full network training cycle and is therefore computationally expensive, BO replaces direct exploration of the search space with a surrogate model based on a Gaussian Process (GP) [53,54]:

$$f(x) \sim \mathcal{GP}\big(m(x), K(x, x')\big) \tag{A13}$$

where the mean function $m(x)$ is set to zero and the covariance function adopts the Matérn–5/2 kernel with Automatic Relevance Determination (ARD):

$$\mathrm{K}_{\nu=5/2}(x, x') = \alpha^2 \left(1 + \sqrt{5}\, r + \frac{5}{3} r^2\right) \exp\left(-\sqrt{5}\, r\right), \quad r = \sqrt{\sum_{d=1}^{D} \frac{(x_d - x'_d)^2}{\ell_d^2}} \tag{A14}$$

where the vectors x and x' represent two candidate hyperparameter configurations in the D-dimensional search space $\mathcal{X}$. The index $d = 1, \ldots, D$ identifies the individual hyperparameters, each associated with a characteristic length-scale ℓ_d, which controls the sensitivity of the surrogate model to variations along that dimension. The quantity r is the ARD-weighted distance between x and x'. Moreover, α denotes the signal variance of the Gaussian Process prior.

After each BO iteration, the GP posterior is updated and the next hyperparameter vector is selected by maximizing the Expected Improvement (EI), which quantifies the probability-weighted gain relative to the best observed performance $f^\star$:

$$\mathrm{EI}(x) = \mathbb{E}[\max(f^\star - f(x),\, 0)] = \sigma(x)[z\, \Phi(z) + \phi(z)], \quad z = \frac{f^\star - \mu(x)}{\sigma(x)} \tag{A15}$$

with $\mu(x)$ and $\sigma(x)$ denoting the GP posterior mean and standard deviation. Φ and ϕ are the standard Gaussian cumulative distribution and density functions.

The architectural, regularization, and training-related hyperparameters are contained in the search space $\mathcal{X}$. The bounds related to each of the previously mentioned hyperparameters have been selected to ensure the computational efficiency of the developed LSTM AE, maintaining a simple structure to avoid prohibitive training times. The neural network model is purposely restricted by the chosen bounds, ensuring both acceptable dimensionality and flexibility for the Bayesian Optimization procedure, which grants sufficient adaptability of the neural network capacity to the multichannel vibration data.

Moderate encoder dimensions are allowed to capture the complex spatial and temporal patterns of the 22-channel vibration data, whereas dropout probabilities and

ℓ_2-regularization coefficients are permitted to vary over broad intervals to effectively prevent overfitting. Logarithmic scaling is adopted for those parameters known to span several orders of magnitude, such as the initial learning rate and regularization strength. The resulting hyperparameter domain is summarized in Table A4.

Table A4. Hyperparameter search space used in the Bayesian Optimization.

Hyperparameter	Range	Type
Encoder units	16–128	even integers
Bottleneck units	0.5· Encoder units	even integers
Decoder units	Encoder units	even integers
Dropout 1	0.1–0.3	Real
Dropout 2	0.1–0.3	Real
Initial learning rate	(10^{-4}–5×10^{-3}) (log scale)	Real
ℓ_2-regularization	1×10^{-6}–1×10^{-3}	Real
Mini-batch size	8–16	Integer

The selected ranges reflect a compromise between model expressiveness, regularization capability, numerical stability, and computational feasibility, all of which are critical when training recurrent architectures on long multichannel vibration sequences. Broader domains were intentionally adopted for those hyperparameters that exert a strong nonlinear influence on the training dynamics, such as the learning rate, dropout probabilities, and ℓ_2-regularization, thereby allowing the optimization process to explore qualitatively different convergence behaviors and regularization regimes.

More restricted intervals have been imposed on parameters whose excessive values may lead to unstable gradients, poor generalization, or prohibitive training times, such as the encoder width and mini-batch size. The BO is executed for eighty iterations.

References

1. Frøseth, G.T.; Rönnquist, A. Evolution of load conditions in the Norwegian railway network and imprecision of historic railway load data. *Struct. Infrastruct. Eng.* **2019**, *15*, 152–169. [CrossRef]
2. Bono, F.M.; Argentino, A.; Bernardini, L.; Benedetti, L.; Cazzulani, G.; Somaschini, C.; Belloli, M. Automated operational modal analysis of a steel truss railway bridge employing free decay response. *J. Infrastruct. Intell. Resil.* **2025**, *4*, 100145. [CrossRef]
3. Argentino, A.; Radicioni, L.; Bono, F.M.; Bernardini, L.; Benedetti, L.; Cazzulani, G.; Belloli, M. Data normalization for the continuous monitoring of a steel truss bridge: A case study from the Italian railway line. *J. Infrastruct. Intell. Resil.* **2025**, *4*, 100171. [CrossRef]
4. Bono, F.M.; Bernardini, L.; Argentino, A.; Anghileri, M.; Capacci, L.; Cazzulani, G.; Belloli, M. OMA applied to a reinforced concrete railway bridge through a reduced measurement set-up. In *Proceedings of the International Operational Modal Analysis Conference*; Springer: Cham, Switzerland, 2024; pp. 548–559.
5. Hasani, H.; Freddi, F. Operational modal analysis on bridges: A comprehensive review. *Infrastructures* **2023**, *8*, 172. [CrossRef]
6. Ashmawi, A.; Nguyen, P.; Jawdhari, A. State-of-the-art review of machine learning applications for bridge inspections. *Adv. Struct. Eng.* **2025**. [CrossRef]
7. Iacovino, C.; Turksezer, Z.I.; Giordano, P.F.; Limongelli, M.P. Comparison of bridge inspection policies in terms of data quality. *J. Bridge Eng.* **2022**, *27*, 04021115. [CrossRef]
8. Suwondo, R.; Hidayat, I.; Suangga, M.; Keintjem, M.; Walewangko, J. Dynamic assessment of a railway bridge using operational modal analysis and fast Fourier transform: A comparative study with finite element analysis. *Eng. Technol. Appl. Sci. Res.* **2025**, *15*, 19200–19206. [CrossRef]
9. Yang, X.M.; Yi, T.H.; Qu, C.X.; Li, H.N.; Liu, H. Performance assessment of a high-speed railway bridge through operational modal analysis. *J. Perform. Constr. Facil.* **2021**, *35*, 04021091. [CrossRef]
10. Sun, Q.; Yan, W.J.; Ren, W.X.; Liu, L.L. Application of transmissibility measurements to operational modal analysis of railway, highway, and pedestrian cable-stayed bridges. *Measurement* **2019**, *148*, 106880. [CrossRef]
11. Rainieri, C.; Fabbrocino, G. *Operational Modal Analysis of Civil Engineering Structures*; Springer: New York, NY, USA, 2014; Volume 142, p. 143.

12. Magalhães, F.; Cunha, A.; Caetano, E. Vibration based structural health monitoring of an arch bridge: From automated OMA to damage detection. *Mech. Syst. Signal Process.* **2012**, *28*, 212–228. [CrossRef]
13. Cabboi, A.; Gentile, C.; Saisi, A. Vibration-based SHM of a centenary bridge: A comparative study between two different automated OMA techniques. In Proceedings of the 9th International Conference on Structural Dynamics (Eurodyn 2014), Porto, Portugal, 30 June–2 July 2014.
14. Braz, W.; Martins, C.; Ortigão, C.; Couto, A. Railway arc bridge analysis through dynamic tests. *Eur. J. Eng. Technol. Res.* **2023**, *8*, 41–45. [CrossRef]
15. Plachý, T.; Polák, M.; Ryjáček, P. Assessment of an old steel railway bridge using dynamic tests. *Procedia Eng.* **2017**, *199*, 3053–3058. [CrossRef]
16. Ticona Melo, L.R.; Silva, R.S.; Bittencourt, T.N.; Bezerra, L.M. Identification of modal parameters in a scale model for a railway bridge. *Int. J. Struct. Stab. Dyn.* **2016**, *16*, 1550059. [CrossRef]
17. Cardoso, M.; Sampaio, R.A.C.; de Souza, R.M.; Silva, E. Operational modal analysis of road-rail bridge. *MATEC Web Conf.* **2015**, *24*, 09001. [CrossRef]
18. Yang, X.M.; Qu, C.X.; Yi, T.H.; Li, H.N.; Liu, H. Dynamic performance analysis of a high-speed railway bridge under train actions using operational modal parameters. *Int. J. Struct. Stab. Dyn.* **2021**, *21*, 2140007. [CrossRef]
19. Svendsen, B.T.; Øiseth, O.; Frøseth, G.T.; Rønnquist, A. A hybrid structural health monitoring approach for damage detection in steel bridges under simulated environmental conditions using numerical and experimental data. *Struct. Health Monit.* **2023**, *22*, 540–561. [CrossRef]
20. Grunert, G. Data and evaluation model for the description of the static–dynamic interface between trains and railway bridges. *Eng. Struct.* **2022**, *262*, 114335. [CrossRef]
21. Nguyen, N.T.; Bui, H.N.; Matos, J.C.; Dang, S.N. Hybrid deep learning framework for damage detection in urban railway bridges based on linear variable differential transformer data. *Appl. Sci.* **2025**, *15*, 12132. [CrossRef]
22. Sunchikala, S.K.; Mohan, S.C.; Gopala, S.D.; Swetha, P. Structural health monitoring of railway truss bridge under moving train load using decision tree models and residual neural networks. *Discov. Civ. Eng.* **2025**, *2*, 35. [CrossRef]
23. Marasco, G.; Oldani, F.; Chiaia, B.; Ventura, G.; Dominici, F.; Rossi, C.; Vecchi, A. Machine learning approach to the safety assessment of a prestressed concrete railway bridge. *Struct. Infrastruct. Eng.* **2024**, *20*, 566–580. [CrossRef]
24. Chalouhi, E.K.; Gonzalez, I.; Gentile, C.; Karoumi, R. Damage detection in railway bridges using machine learning: Application to a historic structure. *Procedia Eng.* **2017**, *199*, 1931–1936. [CrossRef]
25. Ghiasi, A.; Ng, C.T.; Sheikh, A.H. Damage detection of in-service steel railway bridges using a fine k-nearest neighbor machine learning classifier. *Structures* **2022**, *45*, 1920–1935.
26. Minski, L.; Lopez, R.H. Drive-by damage detection in railway bridges using one-dimensional convolutional neural networks. *Proc. Ibero-Lat. Am. Congr. Comput. Methods Eng.* **2023**, *5*, 05.
27. Armijo, A.; Zamora-Sánchez, D. Integration of railway bridge structural health monitoring into the Internet of Things with a digital twin: A case study. *Sensors* **2024**, *24*, 2115. [CrossRef] [PubMed]
28. Kim, S.Y.; Mukhiddinov, M. Data anomaly detection for structural health monitoring based on a convolutional neural network. *Sensors* **2023**, *23*, 8525. [CrossRef]
29. Le-Xuan, T.; Nguyen-Chi, T.; Bui-Tien, T.; Tran-Ngoc, H. ResUNet4T: A potential deep learning model for damage detection based on a numerical case study of a large-scale bridge using time-series data. *Eng. Struct.* **2025**, *340*, 120668.
30. Tran, M.Q.; Sousa, H.S.; Ngo, T.V.; Nguyen, B.D.; Nguyen, Q.T.; Nguyen, H.X.; Dang, S.N. Structural assessment based on vibration measurement tests combined with an artificial neural network for the steel truss bridge. *Appl. Sci.* **2023**, *13*, 7484. [CrossRef]
31. Fernandes, T.M.; Lopez, R.H.; Ribeiro, D.R.F.; Miguel, L.F.F. Early multi-damage classification in railway bridges using drive-by numerical measurements with piecewise aggregate approximation and convolutional neural networks. *Int. J. Struct. Stab. Dyn.* **2025**. [CrossRef]
32. Bayane, I.; Leander, J.; Karoumi, R. An unsupervised machine learning approach for real-time damage detection in bridges. *Eng. Struct.* **2024**, *308*, 117971.
33. Akintunde, E.; Azam, S.E.; Rageh, A.; Linzell, D.G. Unsupervised machine learning for robust bridge damage detection: Full-scale experimental validation. *Eng. Struct.* **2021**, *249*, 113250. [CrossRef]
34. Daneshvar, M.; Sarmadi, H. Unsupervised learning-based damage assessment of full-scale civil structures under long-term and short-term monitoring. *Eng. Struct.* **2022**, *256*, 114059. [CrossRef]
35. Giglioni, V.; Venanzi, I.; Poggioni, V.; Milani, A.; Ubertini, F. Autoencoders for unsupervised real-time bridge health assessment. *Comput. -Aided Civ. Infrastruct. Eng.* **2022**, *38*, 959–974. [CrossRef]
36. Hurtado, A.; Kaur, K.; Alamdari, M.; Atroshchenko, E.; Chang, K.; Kim, C. Unsupervised learning-based framework for indirect structural health monitoring using adversarial autoencoder. *J. Sound. Vib.* **2023**, *550*, 117598. [CrossRef]

37. Entezami, A.; Sarmadi, H.; Behkamal, B.; De Michele, C. On continuous health monitoring of bridges under serious environmental variability by an innovative multi-task unsupervised learning method. *Struct. Infrastruct. Eng.* **2023**, *20*, 1975–1993. [CrossRef]
38. Nurhaliza, R.A.; Octava, M.Q.H.; Hilmy, F.M.; Farooq, U.; Alfian, G. Application of the outlier detection method for web-based blood glucose level monitoring system. *Bull. Electr. Eng. Inform.* **2024**, *13*, 2809–2816. [CrossRef]
39. Zhao, Y.; Lehman, B.; Ball, R.; Mosesian, J.; de Palma, J.F. Outlier detection rules for fault detection in solar photovoltaic arrays. In *2013 Twenty-Eighth Annual IEEE Applied Power Electronics Conference and Exposition (APEC)*; IEEE: New York, NY, USA, 2013; pp. 2913–2920.
40. Yin, X.; Yang, Y.; Huang, Z.; Yan, W. Bridge damage identification based on LSTM network and contact point response. *Int. J. Struct. Stab. Dyn.* **2025**, *25*, 2450268. [CrossRef]
41. Sarwar, M.Z.; Cantero, D. Deep autoencoder architecture for bridge damage assessment using responses from several vehicles. *Eng. Struct.* **2021**, *246*, 113064. [CrossRef]
42. Spínola Neto, M.; Finotti, R.; Barbosa, F.; Cury, A. Structural damage identification using autoencoders: A comparative study. *Buildings* **2024**, *14*, 2014. [CrossRef]
43. Boccagna, R.; Bottini, M.; Petracca, M.; Amelio, A.; Camata, G. Unsupervised deep learning for structural health monitoring. *Big Data Cogn. Comput.* **2023**, *7*, 99. [CrossRef]
44. Li, X.; Xu, L.; Guo, H.; Yang, L. Application of graph convolutional neural networks combined with single-model decision-making fusion neural networks in structural damage recognition. *Sensors* **2023**, *23*, 9327. [CrossRef]
45. Bruni, S.; Collina, A.; Corradi, R.; Diana, G. Numerical simulation of train-track-bridge dynamic interaction. In *Computational Mechanics*; Tsinghua University Press: Beijing, China, 2004; pp. 237–242.
46. Di Gialleonardo, E.; Braghin, F.; Bruni, S. The influence of track modelling options on the simulation of rail vehicle dynamics. *J. Sound. Vib.* **2012**, *331*, 4246–4258. [CrossRef]
47. Bernardini, L.; Collina, A.; Soldavini, G. Railway bridge runability safety analysis in a vessel collision event. *Vibration* **2024**, *7*, 326–350. [CrossRef]
48. Bernardini, L.; Bono, F.M.; Collina, A. Drive-by damage detection and localization exploiting continuous wavelet transform and multiple sparse autoencoders. *Railw. Eng. Sci.* **2025**, *33*, 721–745. [CrossRef]
49. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; No. 2; MIT Press: Cambridge, MA, USA, 2016; Volume 1.
50. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **2016**, *9*, 1735–1780. [CrossRef]
51. Boniol, P.; Liu, Q.; Huang, M.; Palpanas, T.; Paparrizos, J. Dive into time-series anomaly detection: A decade review. *arXiv* **2024**, arXiv:2412.20512. [CrossRef]
52. Lis, A.; Dworakowski, Z.; Czubak, P. An anomaly detection method for rotating machinery monitoring based on the most representative data. *J. Vibroengineering* **2021**, *23*, 861–876. [CrossRef]
53. Snoek, J.; Larochelle, H.; Adams, R.P. Practical bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Process. Syst.* 2012; ISBN 978-162748003-1. Available online: https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf (accessed on 20 April 2026).
54. Wu, J.; Chen, X.Y.; Zhang, H.; Xiong, L.D.; Lei, H.; Deng, S.H. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* **2019**, *17*, 26–40.