Article:
Giglioni, V., Poole, J. orcid.org/0000-0002-7642-9108, Mills, R. et al. (3 more authors) (2025) Transfer learning in bridge monitoring: Laboratory study on domain adaptation for population-based SHM of multispan continuous girder bridges. Mechanical Systems and Signal Processing, 224. 112151. ISSN 0888-3270

https://doi.org/10.1016/j.ymssp.2024.112151

Mechanical Systems and Signal Processing 224 (2025) 112151

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Full Length Article

Transfer learning in bridge monitoring: Laboratory study on domain adaptation for population-based SHM of multispan continuous girder bridges

Valentina Giglioni [a,*], Jack Poole [b], Robin Mills [b], Ilaria Venanzi [a], Filippo Ubertini [a], Keith Worden [b]

[a] Department of Civil and Environmental Engineering, University of Perugia. Via G. Duranti 93, Perugia 06125, Italy
[b] Dynamic Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, S1 3JD, United Kingdom

ARTICLE INFO

Communicated by E. Chatzi

Keywords:
Bridge monitoring
Population-based SHM
Transfer learning
Domain adaptation
Damage classification
Experimental bridge model

ABSTRACT

The presence of sufficient labelled data associated to various environmental conditions and damage scenarios often represents a challenge for the applicability of supervised-learning methods when dealing with structural health monitoring of real-scale infrastructures. To address this problem, population-based structural health monitoring has been recently proposed as an attractive solution, with the goal to collect information from similar structures and transfer health-state labels across the population. This paper focusses on the use of a feature-based transfer learning method. A machine-learning model is trained with source labelled data in a transformed features space to afterwards classify the unlabelled target dataset of a different bridge. More specifically, a domain adaptation-based methodology proposing two possible strategies, a single-source or a multi-source approach, is described. Given the difficulties in validating these techniques on real and varied datasets from multiple bridges, this paper presents a physical benchmark for population-based structural health monitoring in civil-engineering applications. The transfer between different configurations of a laboratory-scale bridge model, subjected to multiple experimental tests under changing environmental conditions and to the same pseudo-damage scenarios, is investigated. The results of the experimental campaign demonstrate the possibility of effectively exchanging damage labels to perform novelty detection and damage classification across the population via domain adaptation, improve the identification of specific damage classes, as well as to increase model's outcome using a multi-source approach, thus overcoming the limitations of conventional machine learning-based methods. Furthermore, this paper provides an open dataset with the physical benchmark-related data, allowing other researchers to test their own algorithms and address the various transfer learning challenges.

1. Introduction

As vital assets of modern society, bridges play a pivotal role in transportation networks for supporting economic growth and contributing to people's social lives. Therefore, a large portion of Structural Health Monitoring (SHM) applications are devoted to ensure the safety and the effective scheduled maintenance activities of infrastructures [1,2]. The huge interest in this topic is

* Corresponding author.
E-mail addresses: valentina.giglioni@unipg.it (V. Giglioni), jpoole4@sheffield.ac.uk (J. Poole), robin.mills@sheffield.ac.uk (R. Mills), ilaria.venanzi@unipg.it (I. Venanzi), filippo.ubertini@unipg.it (F. Ubertini), k.worden@sheffield.ac.uk (K. Worden).

https://doi.org/10.1016/j.ymssp.2024.112151
Received 25 July 2024; Received in revised form 6 October 2024; Accepted 11 November 2024
Available online 26 November 2024

demonstrated by the significant efforts carried out by the cooperation between bridge owners, public institutions and researchers. As indicated in the official document provided by the Ministry of Infrastructure and Transport in Italy in 2021, a significant amount, equal to 1.1 billion euros, is allocated for the financing of interventions aimed at improving conditions of existing bridges and viaducts safety. In this context, given the challenges posed by ageing infrastructure, there is a pressing need for more advanced and sophisticated monitoring solutions to overcome the limits of traditional visual inspections. Recent developments in computer science and sensor technology contributed to enhancing the spread of vibration-based SHM systems, increasingly involving Artificial Intelligence (AI)-based methods [3,4]. In particular, Machine Learning (ML) has emerged as a powerful tool, offering novel approaches to data analysis, pattern recognition and predictive modelling via data-driven approaches, that can anticipate potential degradation or critical failures and rapidly provide feedback [5,6]. In particular, higher levels of damage assessment beyond anomaly detection can be achieved by employing supervised-learning methods, which require labelled datasets to train the model. Looking at the research world panorama within this context, deep-learning has been widely used to detect cracks via image classification, using Convolutional Neural Networks (CNN) [7]. Similarly, a deep learning model was built by Chen et al. [8] to automatically extract the spatial and temporal domain features in the signals and identify the location of small bridge local variations. In parallel, a wide portion of the literature involves the use of several techniques, such as Recurrent Neural Networks (RNNs) or Support Vector Machines (SVMs) to predict damage classes with time-series SHM monitoring data [9–11]. However, because of the difficulty to acquire bridge data under several health-state conditions, most of the applications in real-world scenarios adopt unsupervised learning. In fact, although supervised methods can provide fairly accurate damage-assessment results, they need a sufficient amount of labelled training data representative of both undamaged and damaged states, a requirement that cannot be generally satisfied when monitoring civil infrastructure. Training data may be obtained from Finite Element Models (FEM), but their use is not straightforward in engineering practice, because of high computational costs and modelling challenges. For these reasons, the use of supervised ML for each bridge to be investigated within a transportation network becomes quite challenging. To overcome the limited data availability, which often limits the practical application of SHM to move beyond anomaly detection, the framework of Population-based Structural Health Monitoring (PBSHM) has been recently proposed as a holistic solution in the SHM research community [12]. The goal is to firstly expand the original dataset by collecting data from a set (i.e. population) of similar structures, and afterwards develop ML strategies for a knowledge transfer-based damage assessment. Sharing information within a population would allow one to exploit valuable knowledge to infer diagnosis of a structure for which labels are missing, contributing to building a solid basis towards multi-asset SHM. PBSHM is particularly advantageous for bridge networks, where structures, despite their differences, share common features and similarities in terms of materials and static schemes.

However, it should be highlighted that typical ML classifiers fail to generalise when applied to the whole population, since ML assumes that both training and testing data are required to belong to the same underlying distribution. This fact means that a predictive model is expected to generalise to those future measurements collected from the same domain. Such an assumption implies that each structure with an installed SHM system should be assessed independently using a case-specific supervised ML algorithm. As a solution, Transfer Learning (TL) discusses how to adapt conventional ML to exploit information from a source structure to enhance diagnosis on an unknown target structure. To have a more complete understanding of the underlying theory, Pan and Yang [13] and Zhuang et al. [14] illustrated the theory of TL and reviewed the research progress in several fields. In the SHM framework, many applications involved deep learning-based TL via fine-tuning, basically consisting in modifying a pre-trained model to predict unknown instances of a different structure. Just to cite a few, these modified neural networks were employed by Pan et al. [15] for transferring information between two real long-span bridges and by Azad et al. [16] for SHM of composite structures in limited-data scenarios using experimental tests. Moreover, Bao et al. [17] integrated physics-based and data-driven methods by generating various training data based on the calibrated FEM, pre-training a deep learning network and transferring its embedded knowledge to the real testing domain, demonstrating the efficiency in the context of vibration-based structural assessment of steel frame structures with bolted connection damage. Similarly, a method combining digital twins and TL was proposed by Teng et al. [18], using a pre-trained CNN from the experimentally-tested structure to the real bridge structure. A different approach, belonging to the inductive TL category, involves multi-task learning, where a collection of related tasks is jointly learned by extracting appropriate shared information across the tasks. The idea is that jointly learning can lead to better generalisation performance than considering each single task independently. Such a theory was adopted by Zhang et al. [19] for guided wave-based integrated health monitoring in an aluminium plate and by Wan et al. [20] for SHM missing data reconstruction within a Gaussian process-based Bayesian approach, potentially extending the applicability to the problems of sensor fault diagnosis, change-point detection, and structural condition classification. Successes in a population-level analysis are also described by Bull et al. [21] via an interpretable hierarchical Bayesian approach using operational fleet data.

Despite good performances, some problems inevitably arise when handling real bridge-monitoring situations because of the difficulty to generate even a small quantity of target labels to train the model. To overcome these constraints, transductive TL may be more feasible since it does not require labels from the target domain. Specifically, recent works have been carried out by focussing on a specific sub-category, known as feature-based TL, which mainly refers to Domain Adaptation (DA) [22] in the context of PBSHM. Such a strategy was adopted by Omori et al. [23] to mitigate the data discrepancies for the same bridge before and after retrofit, allowing one to reuse previously existing knowledge in a long-term damage detection process and by Figueiredo et al. [24] to enhance damage-detection performance using FEMs. DA was also applied to detect damage via unsupervised learning across the Z24 bridge and the KW51 bridge [25], and the same structures were further investigated using a Statistic Alignment (SA)-based DA procedure described by Poole et al. [26]. Beyond that, in order to classify damages on an unknown target structure, the two-step DA methodology presented in Giglioni et al. [27] was validated between the Z24 and the S101 benchmark bridges and between the corresponding FEMs. Furthermore, Li et al. [28] proposed a Domain-Adversarial Neural Network (DANN)-based damage detection

method, in which the displacement response of the numerical and the experimental simply-supported beam were used as the input of the DANN model, yielding high accuracy in identifying and localising damage via label transfer.

Within this context, it is worth highlighting that the development of PBSHM technologies is impeded by the lack of suitable datasets to validate TL methods and by the difficulty to create real and varied datasets from multiple similar structures that have data for the same damage-scenarios. While numerous experiments can be found in the literature in the PBSHM sphere, they typically focus on mechanical-engineering structures [29,30]. To gain insights on the applicability of TL on civil infrastructures, the present paper presents a systematic laboratory investigation which aims to create a population of experimental model bridges. This novel experimental benchmark presents four bridge configurations that are characterised by shifted positions of the piers, different span lengths and different surface layers on the deck. The aim is to conduct experimental tests to build up a first-of-its-kind comprehensive dataset with labelled data under changing environment and different health-state conditions. Precisely, with respect to the aforementioned DA application on bridge data, the strengths of the monitoring campaign rely on the simulation of temperature fluctuations and, above all, in the introduction of the same damage scenarios into each bridge configurations. This benchmark therefore provides a basis for testing transfer learning methods for PBSHM; specifically, methods that attempt to leverage information related to changing temperatures and/or damage labels. The dataset can be used to show transfer between two bridges, as previously presented in Giglioni et al. [27], or in a multi-source method, as presented for the first time in this paper, allowing one to enlarge the field applicability. The features are first statistically aligned in each domain of interest and afterwards fed into kernel-based DA methods. The advantage of proposing such a strategy, which is useful when the available labelled data can be extracted from several bridges, is underlined by the improvement of knowledge transfer performance when training the classifier with data from multiple structures rather than using one single source domain.

As the complexity of bridges continues to increase and the demand for sustainable infrastructure grows, TL will play an increasingly critical role in shaping the future of bridge monitoring and maintenance practices. By harnessing the power of TL, engineers can unlock new insights, optimise resource allocation, and enhance the performance and longevity of critical transportation assets. This paper provides promising results from the proposed approach, demonstrating the potential of transfer learning for bridge monitoring, and underlying at the same time as well the main challenges and open questions that need to be deepen.

The rest of the paper is organised as follows: Section 2 illustrates the main steps of the DA theory and the corresponding methodology; Section 3 describes the laboratory model for PBSHM application; Section 4 presents the results of damage detection and classification after DA and a brief discussion of the related limitations and challenges; Section 5 summarises the conclusions of this work.

2. Transfer learning for population-based SHM

The need to gather a large quantity of labelled data to train a robust classifier that generalises well across different structures is the most evident shortcoming for a wide applicability of ML algorithms and still remains a challenge. Following the PBHSM theory, the flow of information throughout a population would allow one to leverage valuable knowledge to evaluate a monitored bridge (i.e., target domain), for which labels are not available, bringing advancements towards multi-asset SHM. The motivation is that specific labelled data, e.g., data associated to damage cases, may only be available for a single member or a restricted number of them, or be generated by an archetypal physical or numerical model. For this reason, the key point is to investigate if the knowledge obtained from a group of structures can be exploited for the whole population.

To handle the differences in data distributions across the population, TL aims to improve generalisation of ML models, so that they can learn health-state information from a source labelled structure and afterwards diagnose the same damage on an unknown target structure. To better understand what TL represents, a domain $\mathcal{D}$ is defined by a feature space $\mathcal{X}$ and a marginal probability distribution $p(X)$, where $X = \{x_i\}_{i=1}^{N}$ is a finite sample from $\mathcal{X}$, while the associated *task* $\mathcal{T}$ is identified by a label space $\mathcal{Y}$ and a predictive function $f(\cdot)$. Given that, a source and a target domain can be expressed, respectively, as $\mathcal{D}_s = \{x_{s,i}, y_{s,i}\}_{i=1}^{N_s}$ and $\mathcal{D}_t = \{x_{t,i}, y_{t,i}\}_{i=1}^{N_t}$, being $x_{s,i} \in \mathcal{X}_s$, $y_{s,i} \in \mathcal{Y}_s$ and $x_{t,i} \in \mathcal{X}_t$, $y_{t,i} \in \mathcal{Y}_t$. Therefore, TL aims at improving the target predictive function $f_t(\cdot)$ in $\mathcal{T}_t$ via the knowledge extracted from $\mathcal{D}_s$ and $\mathcal{T}_s$, assuming $\mathcal{D}_s \neq \mathcal{D}_t$ and/or $\mathcal{T}_s \neq \mathcal{T}_t$. The first assumption results in $\mathcal{X}_s \neq \mathcal{X}_t$ and/or $p(X_s) \neq p(X_t)$, while the second one translates into $\mathcal{Y}_s \neq \mathcal{Y}_t$ and/or $p(Y_s|X_s) \neq p(Y_t|X_t)$. Conversely, a classic ML problem implies that the two domains and their related tasks are the same, i.e., $\mathcal{D}_s = \mathcal{D}_t$ and $\mathcal{T}_s = \mathcal{T}_t$.

TL methods can be categorised according to several criteria. From a label-setting point of view, three subsettings can be mentioned, namely transductive, inductive, and unsupervised TL. The complete definitions are extensively presented in Refs. [13,14]. This paper focusses on transductive TL, assuming the availability of labelled data only from a single source domain, a situation that may reproduce real SHM scenarios. Precisely, a feature-based TL approach is addressed, whose goal is to reduce the gap between feature spaces of source and target domains in terms of proper statistical distances. This idea translates into the application of DA, a branch of TL that aims at minimising the existing gap between marginal distributions $p(X_s) \neq p(X_t)$, or conditional distributions $p(Y_s|X_s) \neq p(Y_t|X_t)$, or both, but assuming that $\mathcal{X}_s = \mathcal{X}_t$ and $\mathcal{Y}_s = \mathcal{Y}_t$.

2.1. The DA-based methodology

In the realm of DA, the flowchart of Fig. 1 represents the adopted methodology for knowledge transfer within a population of bridges, which integrates previous approaches [26,27,30] by providing the possibility to carry out single or multi-source DA, thereby including the case in which labelled data stem from one or multiple structures.

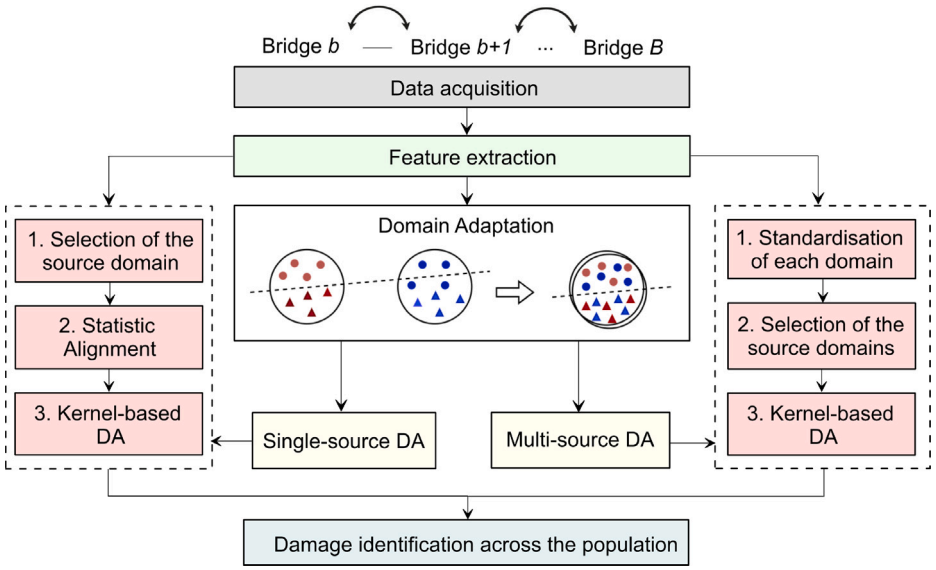

Fig. 1. Flowchart of the adopted methodology for knowledge transfer across B bridges: a single-source or a multi-source DA approach can be carried out depending on the availability of labelled domains.

In detail, data are firstly acquired from B bridges via long-term continuous monitoring and natural frequencies are extracted as damage-sensitive features. Such a population inevitably presents differences in data distribution; therefore, DA is applied to reduce the domain gaps by transforming the features into a common latent space. Two TL situations are described in the methodology, depending on the number of available sources. On the one hand, single-source DA includes the application of SA to firstly align target data with the single pre-defined source domain, assuming the availability of health-state labels. Kernel-based DA, such as Joint Domain Adaptation (JDA), is then applied to further improve the mapping without requiring any target label from damage conditions. On the other hand, when dealing with the transfer across several bridges, the features of each domain of interest are separately standardised as a first step in the multi-source DA approach. The number of sources and target domains is then selected and the corresponding features are therefore fed into JDA, representing in this second case the only step of DA.

In detail, once the two domains in the single-source DA are defined, the alignment is carried out with a two-step procedure. SA is firstly employed to match the first-and-second-order statistics and project the features into a physically-interpretable space. It is demonstrated to be useful in SHM applications, being able to face class imbalance (especially with the scarcity of damage data) and partial DA states [26]. Specifically, Normal Condition Alignment (NCA), introduced by Poole et al. [26], is here adopted to firstly align a limited amount of data generated during normal conditions. This is a typical assumption that could be done in practice, given that the data gathered at the beginning of the monitoring campaign are conceivably associated to the undamaged state. The source domain is transformed via,

$$\overline{x}_s^{(i)} = \frac{x_s^{(i)} - \mu_s}{\sigma_s} \tag{1}$$

with the mean and standard deviation of the feature indicated as μ_s and σ_s, respectively. After computing such quantities during normal conditions, i.e., $\mu_{s,n}$, $\mu_{t,n}$ and $\sigma_{s,n}$, $\sigma_{t,n}$, the new target features are obtained via Eq. (2), that aligns the normal conditions of the target with those of the source,

$$\overline{x}_t^{(i)} = \left(\frac{x_t^{(i)} - \mu_{t,n}}{\sigma_{t,n}}\right)\sigma_{s,n} + \mu_{s,n} \tag{2}$$

To refine the transfer, the second step is carried out by implementing JDA, whose input is represented by $\overline{X}_s$ and $\overline{X}_t$, where $\overline{X}_s = \{\overline{x}_s^{(i)}\}_{i=1}^{N_s}$ and $\overline{X}_t = \{\overline{x}_t^{(i)}\}_{i=1}^{N_t}$. The aim is to find a non-linear transformation $\phi : \overline{\mathcal{X}} \to \mathcal{H}$ that minimises the distance between the joint distributions $p(\overline{Y}_s, \overline{X}_s)$ and $p(\overline{Y}_t, \overline{X}_t)$. Since the joint probability distribution is defined as the product of the marginal and conditional distributions, the problem translates into $p(\phi(\overline{X}_s)) \approx p(\phi(\overline{X}_t))$ and $p(\overline{Y}_s|\phi(\overline{X}_s)) \approx p(\overline{Y}_t|\phi(\overline{X}_t))$. However, given that the conditional distributions are challenging to compute because of the missing target labels, JDA uses class-conditional distributions with a pseudo-labelling approach, which provides the estimates of the labels after training a classifier on the source data. With these assumptions, JDA matches the conditional distributions for each class $p(\phi(\overline{X}_s)|\overline{Y}_s = c)$ and $p(\phi(\overline{X}_t)|\overline{Y}_t = c)$, where $c \in \{1, \dots, C\}$ is in the label set. The class-conditional distributions are therefore determined by optimising the mapping function until the convergence of the target pseudo-labels.

The Maximum Mean Discrepancy (MMD) distance is adopted as a cost function introducing a kernel $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)$. Using the low-rank empirical kernel embedding $\tilde{K} = KWW^TK$, where $K = k(\overline{X}, \overline{X}) \in \mathbb{R}^{(N_s+N_t)\times(N_s+N_t)}$, given $\overline{X} = \overline{X}_s \cup \overline{X}_t \in \mathbb{R}^{(N_s+N_t)\times d}$ and d is the dimension of the features, such a distance can be expressed in terms of a set of weights $W \in \mathbb{R}^{(N_s+N_t)\times k}$, yielding:

$$\mathrm{Dist}(p(\phi(\overline{X}_s)), p(\phi(\overline{X}_t))) + \mathrm{Dist}(p(\phi(\overline{X}_s)|\overline{Y}_s), p(\phi(\overline{X}_t)|\overline{Y}_t)) \approx \mathrm{tr}(W^T K M_c K W) \tag{3}$$

where M_c is the MMD matrix including class labels. The summation in Eq. (3) should be minimised to find the optimal latent mapping, thereby formulating the problem in an optimisation framework, subjected to regularisation, where μ indicates the level of regularisation and kernel PCA removes the trivial solution $W = 0$,

$$\min_{W^T K H K W = I} = \sum_{c=0}^{C} \mathrm{tr}(W^T K M_c K W) + \mu \mathrm{tr}(W^T W) \tag{4}$$

where $H = I - 1/(N_s + N_t)\mathbf{1}$ is the centring matrix with the identity matrix I and a matrix of ones $\mathbf{1}$. The Lagrangian approach is then utilised to work with an eigenvalue problem, where the optimal weights W are obtained from the eigenvectors corresponding to the k smallest eigenvalues from,

$$\left(K \sum_{c=0}^{C} M_c K + \mu I\right) W = K H K W \phi \tag{5}$$

At the end of the process, the k-dimensional transformed feature space is obtained by $Z = KW \in \mathbb{R}^{(N_s+N_t)\times k}$ and a general classifier can be trained and tested in the new latent space containing both source and target data. A detailed description of the method can be found in Refs. [22,31].

Regarding the multi-source DA approach, the methodology suggests to firstly standardise each domain, where centring and scaling are applied independently on each feature by computing the relevant statistics on the samples in the training set, that is collected during undamaged conditions. The transformed features of the bth domain are calculated via Eq. (6), where the means and standard deviations of normal measurements (e.g. undamaged) are stored to be used for the standardisation of the whole domain-referred dataset

$$z_b^{(i)} = \frac{x_b^{(i)} - \mu_{b,n}}{\sigma_{b,n}} \tag{6}$$

The processed features from Eq. (6), extracted from those bridges for which data are labelled, are therefore assigned to the source domain, while the features associated to the unlabelled dataset describe the target domain. After defining each domain of interest, JDA is applied to align the distribution of the target features given a multi-source distribution. Such an approach may become useful when the information extracted from a single source domain is not sufficient for ensuring good TL performance.

Since DA poses the basis to adapt ML algorithms to generalise across a population, a classifier can be implemented to discriminate health-state classes in the target bridge by training the model just using source transformed data. Given that DA should minimise the distance between domains, generating well-concentrated clusters with relatively-close data points, the K-Nearest Neighbours (KNN) algorithm [32] is here adopted for validation, since it is based on the concept of proximity. If DA successfully aligns the feature distributions of source and target domains, the data should be close in the Euclidean space and thus the use of a distance metric becomes meaningful. Known that KNN predictions are based on the distance existing between instances, this algorithm is adopted for the purpose of validating the DA results. In fact, the method assigns to a certain data point the most frequent label among the identified K nearest neighbours, where K is generally recommended to be low (in this paper, a K value equal to 1 or 2 is considered for damage-identification purposes). Note also that when the domains' shift is minimised, KNN's tendency to favour the majority class is mitigated. Moreover, since the focus here is on DA, KNN could be considered as an indicator of the alignment performance. However, it should be underlined that this is not necessarily the classifier to be used in practice. Any ML technique can be selected after DA, given that real scenarios may require the application of more sophisticated algorithms.

3. Description of the laboratory setup

The experimental model bridge is located inside the structural dynamics Laboratory for Verification and Validation (LVV) of the University of Sheffield. The structure is aluminium and is placed on a 3 m × 2 m shaker table to reproduce random input vibrations with a frequency range of 5–110 Hz (Fig. 2). The main deck, with a global length of 2990 mm, a thickness of 2 mm and a width of 270 mm, is made of four continuous "I" beams (Fig. 2 e), that are bonded to a horizontal top plate with Scotch-Weld Acrylic Adhesive DP8425NS. The piers, characterised by Bosch-Rexroth sections, are clamped to the MAST surface with two steel plates, which are attached together with Belleville washers to allow for variable connection stiffness between the MAST surface and the pier.

Regarding the modelling of the joints deck/piers, flanged deep-groove ball bearings are utilised to restrict lateral movements. The use of such modular bearings, clamped to the "I"-beam section, allows either pinned conditions (i.e. allowing rotation only, no axial or vertical motion) or rolling-pinned conditions (i.e. allowing rotation and axial movement, but no vertical movement). Precisely, while the longitudinal movement is prevented in just one of the end piers, that is pier 4 (Fig. 4 a), the intermediate ones and un-fixed end are free to roll axially to allow for thermal expansion and motion resulting from any induced vibration (Fig. 4 b). The connection system permits to move the piers anywhere along the axial length of the bridge deck, translating into the possibility to reproduce any required span length.

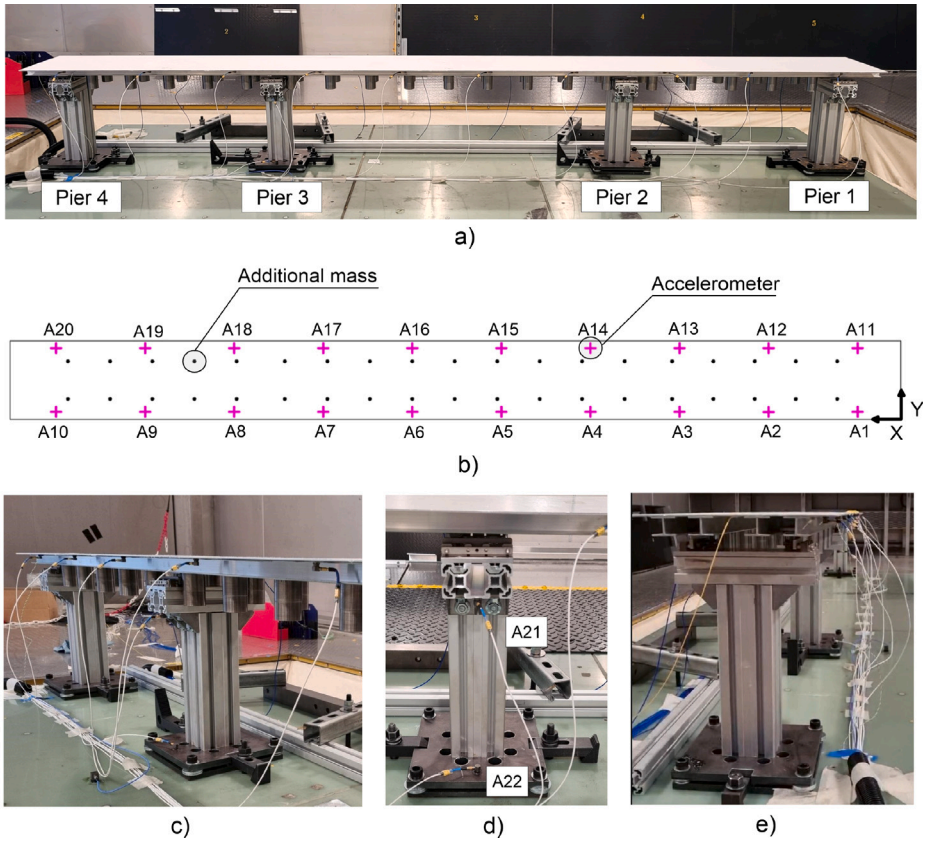

Fig. 2. General view of the mockup bridge (a), scheme of the bridge deck, including the position of the uni-axial accelerometers and the additional masses (b), detailed views of the bridge structure (c), sensors location at the pier (d), deck section (e).

The global weight of the deck is ~10 kg; however, given the high spans stiffnesses, additional masses are homogeneously distributed along the deck with the aim to increase the weight to ~ 30 kg and thus reduce the natural frequencies. The positions of these masses is visible in the scheme of Fig. 2 (a), which also shows the location of the accelerometers, named from A1 to A22, that acquire measurements along the vertical direction. Their coordinates and sensitivity values are listed in Table 1. Note that the adopted reference system has the Z-axis pointing downwards and, moreover, most of sensors, i.e., A1–A20, are situated on the deck, while the remaining A21 and A22 sensors are placed along pier 3, as depicted in Fig. 2 (d). Specifically, data acquisition consists in collecting consecutive ~ 20 second-long repeats at a sampling frequency of 256 Hz.

The monitoring system is integrated with a thermocouple on the deck to measure temperature during the experimental tests. In particular, to simulate various operational and environmental conditions, a wide range of temperature is generated within the environmental chamber, from −15 °C to 30 °C. Four different bridge configurations, named here as B1, B2, B3 and B4, are built with variation introduced by changing the number and the position of the piers and the surface layer on the deck. Table 2 describes the corresponding geometries. B1 is a three-span bridge where the main span measures 1.29 m and the deck is covered by transparent adhesive tape. B2 and B4 configurations present the same number of piers, with the main spans equal to 1.37 m and 1.55 m, respectively, and they are both characterised by an additional woven cotton sheet on the top of the deck and lubricant on the bearings. This new layer, attached using a heat fusible adhesive, is introduced to create a freezable water layer during the lowering of the temperature. The removal of the Pier 2 from the previous B2 configuration yields a two-span bridge, indicated as B3, with a main span of 2.045 m. A summary of the differences among the population's members is presented in Table 3.

3.1. Simulations of health-state scenarios

Each configuration is subjected to several damage scenarios and environmental conditions, with the aim to build a comprehensive dataset and study the structural response in terms of modal properties variation. Damage classes can be broadly divided into two macro-categories, the former describing a pseudo-stiffness reduction caused by the maximum bending moment, while the latter involves the seizing of the bearings. Each category includes different typologies of damage data that are continuously acquired at ambient temperature. By applying a 64 g mass, M1 and M2 scenarios simulate decays in natural frequencies that can be considered as equivalent to those caused by a stiffness reduction on the side and centre line, respectively, at $L_{lat}/2$, where L_{lat} is the length

Table 1
Location of the accelerometers and sensitivity values.

Sensor	Sensor location (X,Y,Z) [m]	Sensitivity [mV/g]
A1	(0.145,0.025,0)	98.8
A2	(0.445,0.025,0)	98.7
A3	(0.745,0.025,0)	98.7
A4	(1.045,0.025,0)	100.6
A5	(1.345,0.025,0)	101
A6	(1.645,0.025,0)	97.3
A7	(1.945,0.025,0)	101.6
A8	(2.245,0.025,0)	100.7
A9	(2.545,0.025,0)	98.4
A10	(2.845,0.025,0)	98.2
A11	(0.145,0.245,0)	97.1
A12	(0.445,0.245,0)	96.4
A13	(0.745,0.245,0)	99.7
A14	(1.045,0.245,0)	99.2
A15	(1.345,0.245,0)	100.1
A16	(1.645,0.245,0)	100.9
A17	(1.945,0.245,0)	91.6
A18	(2.245,0.245,0)	100
A19	(2.545,0.245,0)	98.1
A20	(2.845,0.245,0)	98.4
A21	(2.135,0.045,−0.095)	100
A22	(2.135,0.055,−0.355)	100

Table 2
Position of the piers for the four bridge configurations.

Bridge name	Deck length [m]	Pier 1 [m]	Pier 2 [m]	Pier 3 [m]	Pier 4 [m]
B1	2.99	0.14	0.855	2.145	2.86
B2	2.99	0.14	0.815	2.185	2.86
B3	2.99	0.14	–	2.185	2.86
B4	2.99	0.14	0.725	2.275	2.86

Table 3
Differences between bridge configurations.

Bridge name	≠ Pier number	≠ Spans length	≠ Surface layer
B1	–	✓	✓
B2	–	✓	–
B3	✓	✓	–
B4	–	✓	–

of the lateral span. The same mass is placed on the side and centre line at $L_{main}/2$, where L_{main} is the length of the main span, to generate M3 and M4 classes, respectively. It is important to note that using added masses, instead of cuts or physical damage, to induce changes in the dynamic behaviour of the mockup allowed for complete reversibility of the different scenarios, thereby preventing damage accumulation. A scheme with the selected damage locations is depicted in Fig. 3 (b), while illustrative pictures of M3 and M1 damage conditions are visible in Fig. 3 (a) and (c).

Low-severity versions of the scenarios M1–M4 are reproduced by using a mass of 21.6 g, yielding M5–M8 damage labels.

Moving to the second damage macro-category, SB1 and SB2 denote the situation in which the bearing of Pier 1 is seized by preventing respectively longitudinal movements and both longitudinal and rotational movements (Fig. 4 c).

In addition, the longitudinal direction along the axial length of the deck is locked for Pier 2 (SB3) and for both Pier 1 and Pier 2 (SB4). A brief description of the simulated scenarios is listed in Table 4. It should be underlined that all the induced structural states are independently applied and are not affected by the previous experiments. As reported in Table 5, the single configurations are characterised by a variable number of health-state classes, and each class includes a different number of data points. Note that the undamaged-state data labelled with "N" collected from each structure represent the most-populated category, especially for B2 and B4. In fact, while B1 and B3 contain vibration measurements at ambient temperature, healthy accelerations of B2 and B4 are gained under changing environmental conditions. Specifically, the environmental chamber is set to reproduce temperatures in a large range, equal to −5 °C–15 °C and −15 °C–15 °C during the acquisition of B2 and B4, respectively. The same range of temperatures is then adopted to collect data for the waterlogged scenario (W), in which the woven cotton sheet over the deck surface is sprayed with water to simulate freezing effects. Therefore, the dynamic behaviour of B2 and B4 is affected by a significant temperature variation over the whole monitoring campaign.

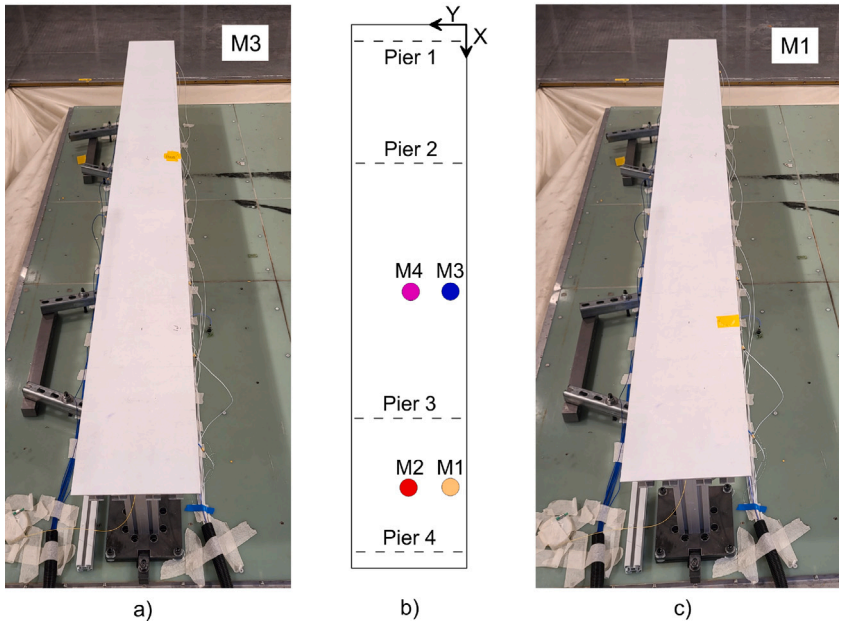

Fig. 3. Illustrative pictures of damage scenarios M3 (a) and M1 (c). The scheme of the selected damage locations is shown in (b).

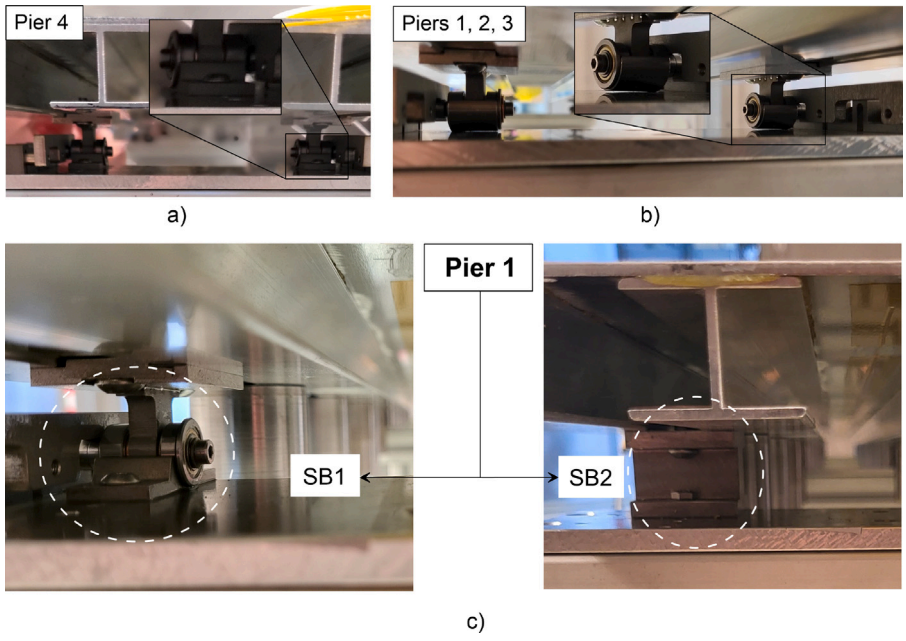

Fig. 4. Longitudinal movement is locked at pier 4 (a) and allowed for piers 1, 2 and 3 (b). Seizing of the bearings at pier 1: SB1 and SB2 damage scenarios (c).

3.2. Feature extraction

After signal pre-processing, automated system identification is carried out for each bridge configuration to extract modal parameters via the covariance-based Stochastic Subspace Identification (SSI) technique within the MOSS environment [33]. To handle consistent feature spaces for DA, the four mode shapes illustrated in Fig. 5 are selected, showing different values of natural frequencies but similar characteristics.

The first, second and third bending modes are indicated as Mode 1, Mode 3 and Mode 4, while Mode 2 represents the torsional mode, where the torsion largely affects the main span. A comparison between them is made in terms of Cross Modal Assurance Criterion (CMAC) values, computed by analysing each pair of configurations (Table 6). While a good correlation, especially for the

Table 4
A brief description of the simulated scenarios.

Label	Description
N	Undamaged conditions
W	Waterlogged deck
M1	Mass of 64 g - lateral span (on the side)
M2	Mass of 64 g - lateral span (in the middle)
M3	Mass of 64 g - main span (on the side)
M4	Mass of 64 g - main span (in the middle)
M5	Mass of 21.6 g - lateral span (on the side)
M6	Mass of 21.6 g - lateral span (in the middle)
M7	Mass of 21.6 g - main span (on the side)
M8	Mass of 21.6 g - main span (in the middle)
SB1	Seizing of the bearing at pier 1 (locked longitudinal movement)
SB2	Seizing of the bearing at pier 1 (locked longitudinal and rotational movements)
SB3	Seizing of the bearing at pier 2 (locked longitudinal movements)
SB4	Seizing of the bearing at pier 1 and pier 2 (locked longitudinal movements)

Table 5
Number of data points for each health-state class varying the bridge configuration.

Bridge	N	W	M1	M2	M3	M4	M5	M6	M7	M8	SB1	SB2	SB3	SB4
B1	36	–	10	10	10	10	5	5	5	5	–	–	–	–
B2	125	73	10	10	10	10	10	5	5	5	10	10	10	–
B3	20	–	10	10	10	10	–	–	–	–	–	–	–	–
B4	200	121	10	10	10	10	10	10	10	10	10	–	10	10

Table 6
CMAC values for each pair of configurations.

Mode	B1-B2	B2-B3	B1-B3	B2-B4	B3-B4	B1-B4
Mode 1	0.99	0.33	0.39	0.96	0.46	0.99
Mode 2	0.99	0.54	0.57	0.98	0.64	0.99
Mode 3	0.93	0.07	0.04	0.5	0.01	0.59
Mode 4	0.47	0.13	0.17	0.63	0.36	0.78

Table 7
Reference natural frequencies of each bridge configuration.

Bridge name	F1 [Hz]	F2 [Hz]	F3 [Hz]	F4 [Hz]
B1	37.67	58.70	85.37	92.41
B2	41.31	62.82	80.65	85.79
B3	14.78	41.33	47.53	81.77
B4	33.26	54.73	88.43	103.01

first modes, can be found between B1, B2 and B4 configurations, which hold similarities in terms of number of spans, the comparison of such configurations with B3, that has one less span, reasonably yields relatively-low CMAC values. If the interest is to extract mode-shapes-based damage sensitive features, which is not the case of this work, a comparison by isolating the damaged span is thus recommended. A study on the relation between CMAC and transfer outcomes is investigated in Poole et al. [34].

The reference frequencies corresponding to each mode, i.e., F1, F2, F3 and F4, listed in Table 7, are obtained from a single acquisition during undamaged conditions and are afterwards adopted for automated frequency tracking over the whole monitoring period.

The details on signal processing and frequency tracking are omitted for the sake of brevity, given that they are not the focus of this paper.

By looking at feature (i.e., the selected natural frequencies) distributions, the influence of environmental variations on those configurations subjected to a wide range of temperatures are clear. Such an effect is appreciable in Fig. 6 (c), where the first frequency of B2 is plotted against temperature, which is decreased from 15 °C to −5 °C and afterwards brought back to the starting value. Despite the shift between normal (N) and waterlogged (W) clusters, the trend of F1 is the same in both cases, exhibiting a reasonable growth with low temperatures to reflect a higher stiffness. The same conclusion comes from Fig. 6 (d), showing the tracking of F1 under the aforementioned range of simulated temperatures. Note that the clear peaks at the top of the two bell curves indicate the stiffness increase in correspondence to the minimum temperature value registered during normal (N) and waterlogged (W) scenarios, respectively, where the waterlogged natural frequencies are smaller because of the effect of the added mass of the water.

Regarding the damage effects on the selected features, it is possible to highlight different frequency variations that are consistent with the type of damage. As an example, Figs. 6 (a) and (b) show that the use of F2 and F3 is discriminative for damage classification

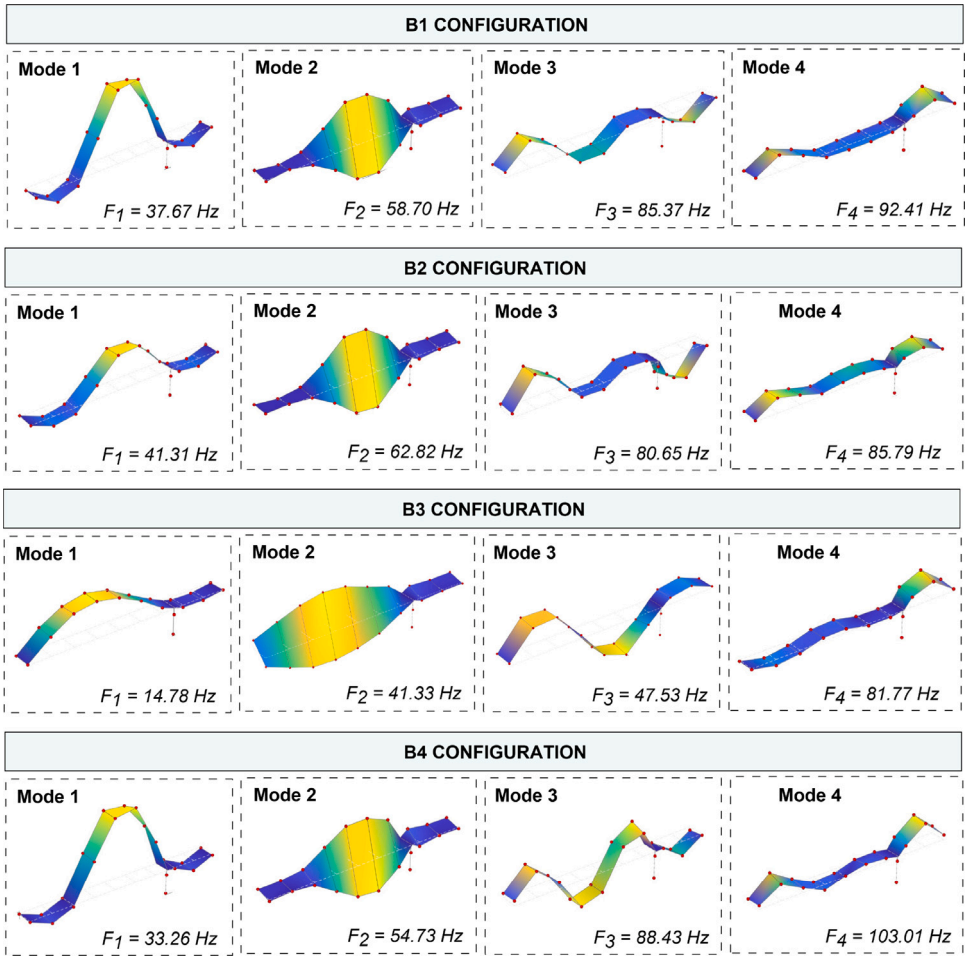

Fig. 5. Identified mode shapes for each bridge configuration extracted from the post-processing of a single sample during healthy conditions.

in B1. The second frequency significantly contributes to visualising the effect induced at the main span by the mass on one side (M3) and the midline of the deck (M4). In fact, F2 represents the mode mainly affecting the same span on which the mass is located. In particular, being a torsional mode, more evident decays are associated with damage M3 (Fig. 6 a). Conversely, the opposite situation occurs when looking at Fig. 6 (b), as F3, the frequency of the mode involving the lateral spans, is more discriminative for M1 and M2 (the mass is on the lateral span).

An example of damage classification on B2 using the KNN is provided in Fig. 7. Although the confusion matrix demonstrates a good performance, it is worth pointing out that some classes cannot be perfectly identified and that each class is populated by a restricted number of instances, translating into a lack of a sufficiently-large dataset. Such an issue may negatively affect the generalisation process and the robustness of the supervised classifier on the single bridge configuration. As a possible solution, the collection of more labelled data from other similar configurations would allow one to train the model with transformed data of a given labelled structure and afterwards test the performance to classify health-state classes of a different bridge.

4. DA results within the population of experimental model bridges

In order to investigate domain adaptation, various tasks are addressed, trying to simulate a number of scenarios depending on the labels that are assumed available for the configuration of interest. It is specified that the symbols "∘" and "×" are adopted to represent source and target data, respectively, and that healthy data are indicated in the plots with "H". Starting with the single-source approach, the previously-extracted natural frequencies are transformed via NCA and JDA, leading to a common latent space in which the two selected domains can be represented. Here, the KNN is trained on one source domain and tested on a different target domain. The transfer of each single damage class across each pair of configurations is firstly evaluated in terms of F1 score computed with macro average. Such a quantity, combining the True Positive rate (TPr) with the precision, takes into account the relation of the correctly-classified positive samples with both false negatives and false positives and is more reflective of performance when there is class imbalance. Better performance is described by F1 score values equal to 1. The aim to evaluate whether DA can

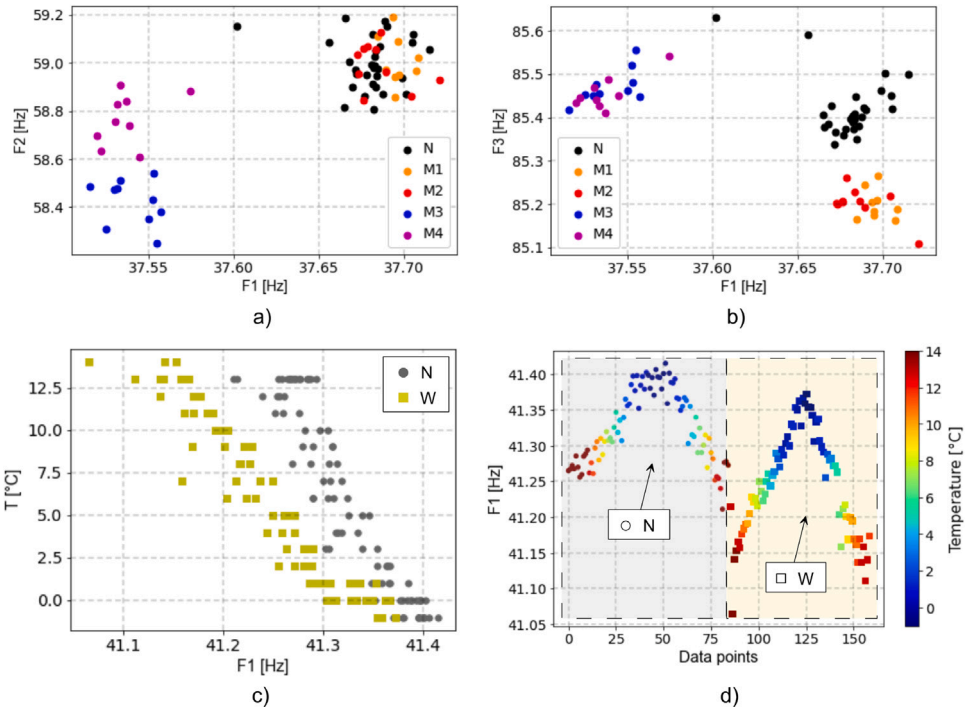

Fig. 6. System identification results of B1 varying health-state scenarios: (a) the relationship between F1 and F2 in B1; (b) the relationship between F1 and F3 in B1. The effect of temperature in B2: the trend of F1 vs temperature (c) and frequency tracking (d) during normal (N) and waterlogged (W) conditions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

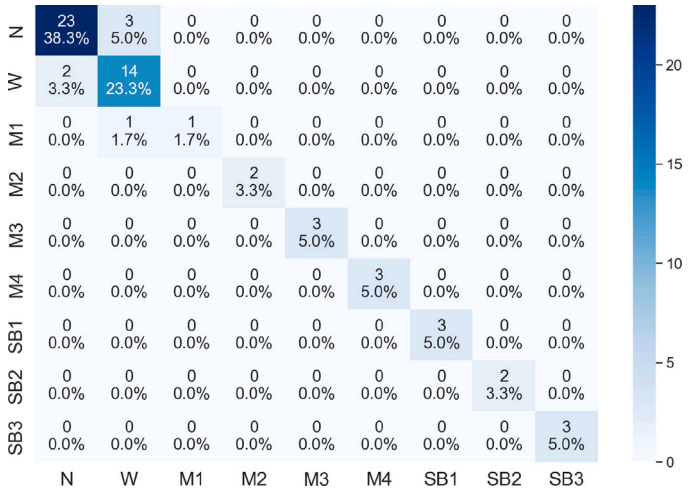

Fig. 7. Confusion matrix after the use of KNN for supervised damage classification in B2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

improve generalisation of a model across bridges where only the deck length is changing, and material and general construction is controlled. The results are reported in Table 8, showing that damage detection performance clearly changes when varying the type of damage scenario to be transferred. It is specified that, given a generic pair X ⟷ Y, each value within the table is a mean between X ⟶ Y and Y ⟶ X. More detailed information is provided in , where all the possible combinations are investigated, since the selection of a certain domain as source or target affects the final outcome, and then evaluated by computing F1 score, TPr and the False Positive rate (FPr). Overall, the DA-based damage detection approach can effectively transfer label information across the population, providing a global F1 score, mediated over all possible cases, equal to 0.8. To avoid misleading observations, a more detailed interpretation is however required. Hence, to better explain the dependence of DA results on damage class (which has a common pattern for all the analyses shown in Tables 8), Fig. 8 represents the three adopted performance metrics associated with

Table 8
F1 score after DA for each pair of configurations.

d	B1 ⟷ B2	B1 ⟷ B3	B2 ⟷ B4	B3 ⟷ B4	B1 ⟷ B4	B2 ⟷ B3
M1	0.73	0.92	0.49	0.56	0.46	0.77
M2	0.84	0.68	0.57	0.83	0.58	0.87
M3	0.88	1	0.99	0.91	0.9	0.96
M4	0.93	1	0.8	0.82	0.9	0.9
M5	0.65	–	0.87	–	0.7	–
M6	0.61	–	0.93	–	0.64	–
M7	0.86	–	0.94	–	0.88	–
M8	0.7	–	1	–	0.84	–
SB1	–	–	0.54	–	–	–
SB3	–	–	0.91	–	–	–

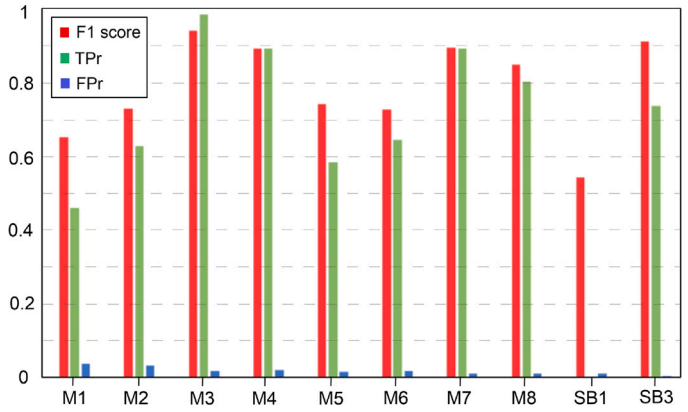

Fig. 8. DA results, in terms of F1 score, TPr and FPr, for each single damage class by considering all the combinations across the population. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the transfer of each single damage class. The rate of false positives keeps a constant and low value throughout the scenarios. On the other hand, note that the capability to correctly recognise damage labels across two different configurations is higher if looking at the damage induced in the main span, i.e., M3 or M4, producing a F1 score and a TPr over 0.9. A reasonable performance decay is visible considering a lower damage severity (M7 or M8). Damage caused by the mass on the lateral span has a lower impact on the selected features and, consequently, might be more difficult to be transferred. This fact reflects the need to correlate TL outcomes with the selection of the most suitable features for the specific kind of damage. An important initial step is in fact to ensure features satisfy the assumption made by DA, i.e., that their response to damage is relate.

Some DA applications on this data set are presented hereafter. Fig. 9 (a) and (b) show the capability to align damage M4 from B2 to B3 and M3 from B4 to B3, respectively. A KNN is trained using labelled data of the source configuration and afterwards tested to identify such a damage in B3, with the F1 score equal to 0.96 and 0.95, respectively. On the other hand, as mentioned before, the classification of the damage simulated in the short span may be quite challenging because of the risk of masking effects. In fact, as inferred from the third row of the confusion matrix of Fig. 7, M1 is not well recognised in the B2 configuration. In this specific case, an additional reason for the poor detection could be the significant lack of labelled data points for training the algorithm, which prevents a robust judgement. An improvement is provided by applying DA and exploiting source labels from B3 (Fig. 9 c), yielding a perfect TPr.

Focussing on the second category of simulated damage, i.e., the seizing of the bearings, Table 8 and the histograms in Fig. 8 show that SB3 labels are successfully transferred between B2 and B4 (being the only two configurations in which this damage is applied). The new shared domain is represented in Fig. 9 (d). Since the healthy measurements related to B2 and B4 are extracted under a changing environment, the distribution of the post-alignment features at different temperature ranges, i.e., $T < 3$ °C, 3 °C $< T \leq 20$ °C, $T > 20$ °C, is represented in Fig. 9 (e). In this case, M3 is considered as an example of damage class. Considering the missing overlapping of data from different temperatures in the original feature space, the successful alignment brought by DA is demonstrated by the fact that source and target instances labelled with "H" belong to the same cluster, despite the wide variability in temperature values. However, it is also possible to notice that, as T[°C] increases, the risk of masking damage is higher. Such a situation can be better visualised by looking at Fig. 10 (a), where the SB1 cluster completely overlaps healthy data, yielding a null TPr. Here, a damage classification problem is addressed, trying to align three different clusters, two of which are describing damage conditions. Although the corresponding labels are correctly recognised, note that certain damages (SB1 in this case) can be difficult to detect; the main reasons include the severity level, the type of feature (and its damage sensitivity) and the temperature fluctuations in B2 and B4, producing a significant impact on data distributions, as shown in Fig. 9 (e). This fact confirm that, prior

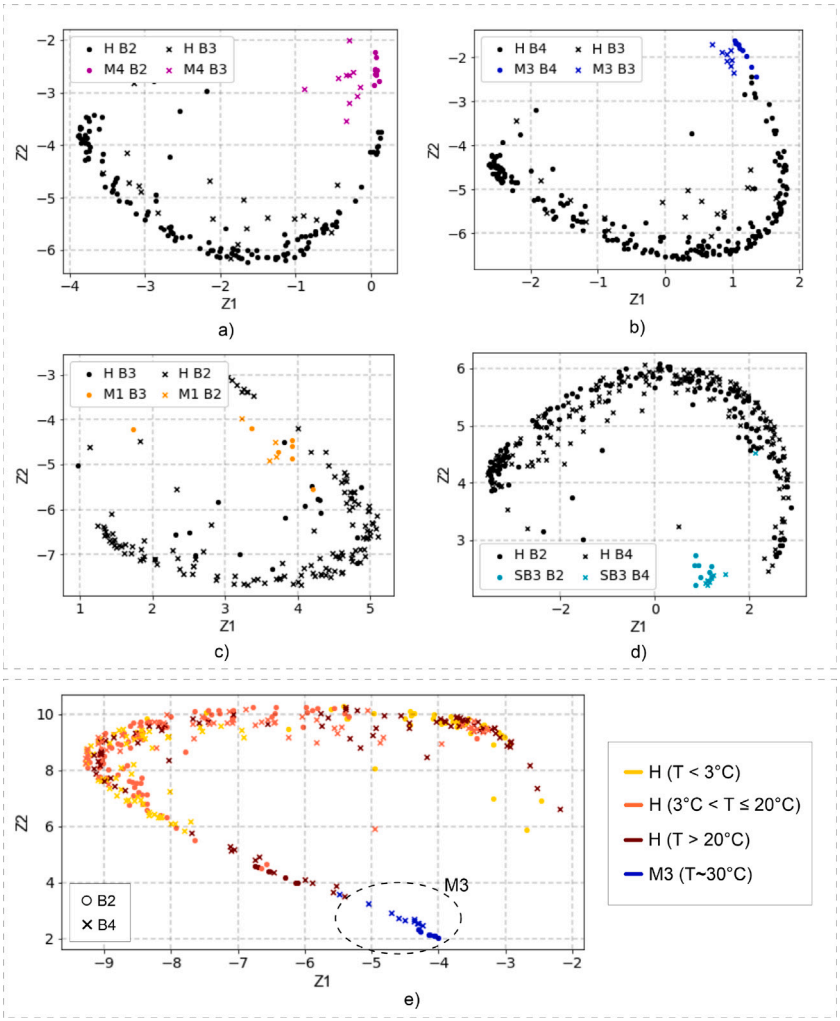

Fig. 9. Damage detection results after DA. The alignment and the transfer of the health-state classes (H)-(M4) and (H)-(M3) between B2-B3 and between B4-B3 is shown in (a) and (b), respectively, where the cluster (H) includes all the data acquired during normal and waterlogged conditions, when simulated. DA is also applied to transfer damage labels associated to M1 (c) and SB3 (d) classes across B3-B2 and B2-B4, respectively. The distribution of the post-alignment features from B2 to B4, belonging to "H" class, is represented in (d) at different temperature ranges. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

to the TL performance evaluation, proper considerations about the possibility to remove environmental effects and to study the relation between the damage to be detected and the most sensitive feature should be made.

Additional damage classification results are depicted in Fig. 10. An interesting application is learning how to correctly classify across the source and target domains, those damages affecting two different spans. The goal is therefore to discriminate between M1/M2 and M3/M4 within the source domain and then exploit this information to identify the same patterns in the target configuration. Figs. 10 (b), (c) and (d) represent DA results aimed at simultaneously transferring label information obtained from the main and lateral spans, yielding F1 scores equal to 0.98, 0.86 and 0.77, respectively. In the last case, although the TPr is 1 for each damage class, a slight increase of the false positives causes a global lower performance.

Despite these good results, it is worth underlining that damage classification is a challenging task, which strictly depends upon the percentage of success in detecting a single damage class in each domain of interest. Therefore, based on those scenarios that are effectively transferred according to , DA can be applied using more than two health-state classes. The analyses are not reported here for the seek of brevity, but some considerations and limitations should be highlighted. While the simultaneous transfer of M1/M2 with M3/M4 generally leads to positive outcomes (see Fig. 10), TL performance is not sufficient (less than a random analysis) when the aim is to transfer low severity damages (M5/M6 with M7/M8) or to recognise across domains the damage simulated at the middle line, e.g. M1, from the one simulated at the side of the span, e.g. M2. It means that there is a huge number of possible tasks, related on case-dependent SHM purposes and depending on single-damage TL behaviour.

Fig. 11 illustrates a similar but slightly different situation with respect to the previous one in Fig. 10 (a), in which a certain damage, M4 in this case, is masked just in one configuration, that is B4. The alignment is therefore negatively affected, resulting

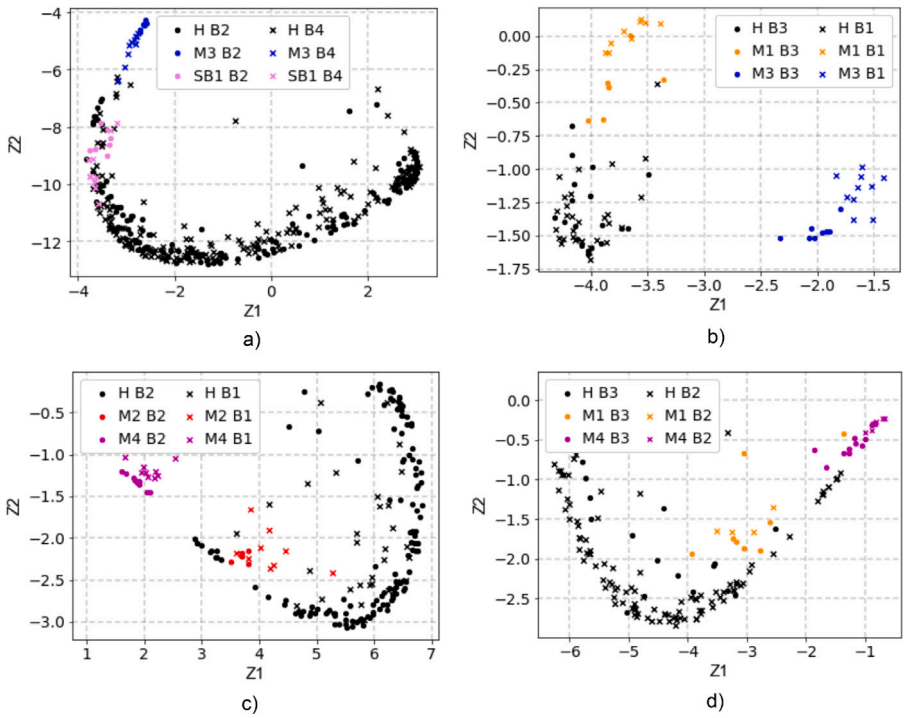

Fig. 10. Damage classification is performed via DA. Health-state classes are well-aligned between B2 and B4, but SB1 damage is totally masked in both configurations (a). The classification of the damage simulated in the main span (M3,M4) from the damage simulated in the lateral span (M1,M2) is carried out across different domains (b, c, d). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

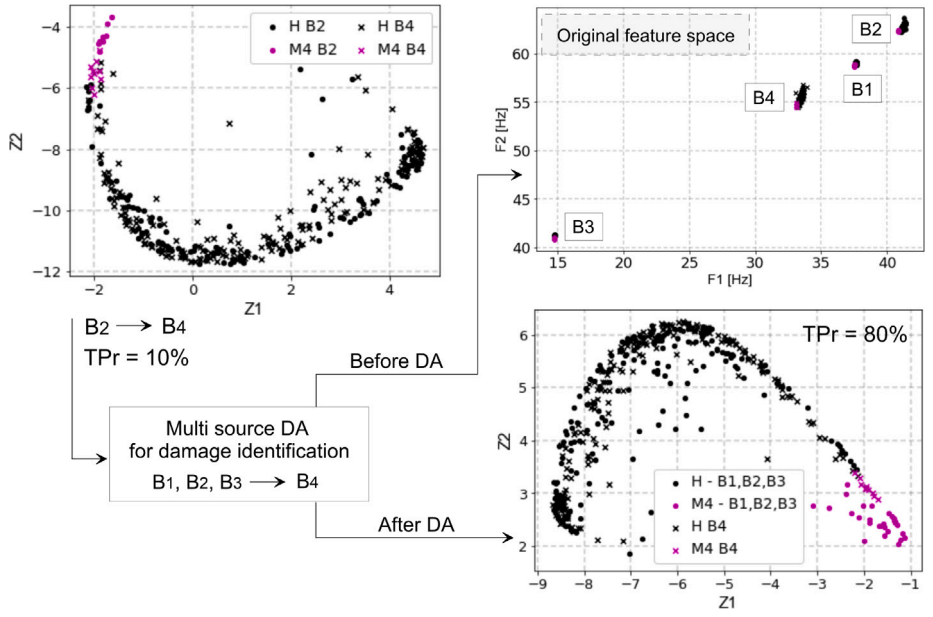

Fig. 11. Multi-source DA results: the information extracted from B1, B2 and B3 is used to classify damage M4 in B4, providing an improvement in the TPr in the new transformed space. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

thus in a low percentage of TPr, i.e., 0.1, when training the ML classifier using B2 data. For this reason, a multi-source DA approach is proposed as a possible help by checking if the use of multiple sources, namely B1, B2 and B3, might improve TL results in terms of false detection errors from B4 label predictions. Plotting all data distributions in the original feature space, it can be noticed that health-state clusters are quite scattered, hampering the effectiveness of a classifier. The necessity to reduce the discrepancy across

Table 9
Multi-source DA-based damage detection results.

Task	Target	Source		F1 score	
		Single	Multi	Single-source DA	Multi-source DA
M8	B1	B2	B2, B4	0.64	0.86
M3	B2	B3	B1, B4, B2	0.91	1
M4	B4	B2	B1, B2, B3	0.64	0.94
M7	B4	B1	B1, B2	0.76	0.93

domains distributions is therefore clear. Such a solution brings important improvements in better discriminating between damage and undamaged conditions. By analysing the KNN outcomes, it is clear that the presence of more sources presenting M4 labels provides an increase in the TPr, rising from 10% to 80%. To better highlight the related advantages, the multi-source method is tested on additional situations characterised by a poor damage detection performance. Examples are summarised in Table 9. The capability to identify the M8 and M4 classes in B1 and B4, respectively, given the knowledge extracted from B2, led to a F1 score equal to 0.64. The introduction of a multi-source domain provides in both cases a substantial improvement, demonstrated by 0.86 and 0.94 F1 score values. Slightly lower yet noticeable benefits are observed during the detection of M3 in B2 and M7 in B4, where the improvement in the performance metric is attributed to a reduction in false-positive rates.

The presented results are quite promising since they demonstrate the real possibility to leverage information between bridges that present same materials, same static scheme, same deck cross-sections and same length and width, but differ in the number and length of the spans. In the context of bridge similarity, several aspects deserve a deeper investigation. The information regarding the damage class of interest could be useful to define a similarity criterion together with the geometry, static scheme, materials and boundary conditions. In general, prior analyses on the correlations between the type of damage to be transferred, the type of feature and the type of structures to investigate are necessary to develop a broader and clearer view on the inner relationship underlying the transfer. Feature selection, which is not the focus of this study, represents a fundamental step prior to DA and, although the natural frequencies are clearly suitable for the paper's purpose, they may not be sensitive to localised damage. It may be therefore recommended to firstly define which is the damage of interest, then select the most discriminative features (e.g. natural frequencies, mode shapes, raw vibration measurements and more) to ensure the possibility to get a good transfer and, as a last step, choose the most suitable TL method. To avoid diverting attention from the focus of this paper, i.e., transfer learning, natural frequencies were used as features, but future work could identify better features for damage identification.

Beyond that, an important challenge to face is how to handle those situations in which two domains experience completely different health-state classes. This is not the aim of this paper, because the hypothesis at the base of DA (i.e., the same number and type of scenarios between the source and the target), do not allow to take into account all the real possible SHM situations. Regarding the damage simulation process, it should be underlined that the main advantage of introducing reversible and easily repeatable scenarios is the possibility to study and validate DA effectiveness for each damage of interest, which is the focus of this work. However, real-world situations are characterised by different degrees of complexity, where damage accumulation and several additional physical phenomena such a nonlinearity, represent challenging situations to deal with. This topic deserves further investigations, with the aim to extend the knowledge on TL to those cases in which more than one damage could occur and accumulate over time. In these cases, DA may not be recommended. Being a feature-based TL method, there is a strict dependence with the adopted features, and such features could behave differently towards multiple damages of different typologies that simultaneously occur, thus negatively affecting the transfer performance.

Handling and removing environmental effects still represents a challenge in the field of TL. Although this work aims to assess the effectiveness of domain alignment under changing environmental conditions, it is necessary to underline that temperature variations can hide the presence of damage, providing a negative impact on anomaly detection results. Therefore, it becomes imperative to understand and investigate the benefits brought by the application of DA in conjunction with the removal of external factors. In this framework, this dataset also provides a way to test methods that use transfer learning/multitask learning for data normalisation and to explore the removal of temperature effects across the population. Particular attention should be directed to the fact that each domain of interest may be subjected to different data cleansing and data normalisation techniques, resulting in possible alterations in the residual distributions.

5. Conclusions

The concept of PBSHM via TL is starting to be applied to bridge SHM, overcoming a significant limitation of traditional data-based SHM approaches, where inferences are limited to novelty detection in the absence of labelled training data. The available dataset can thus be enriched by collecting more information about health-state labels from a population of similar bridges, sharing common features in terms of materials and structural schemes. Dealing with civil infrastructures, DA, belonging to transductive TL methods, is a suitable choice since it does not require target labels. In this light, a DA-based methodology is presented in such a way that two possible strategies could be implemented, namely a single-source or a multi-source approach, based on the number of bridges with a completely labelled dataset. However, the application of such techniques is quite challenging given the difficulties to collect real datasets from multiple similar structures with different health-states. To fill this gap, as well as to validate and discuss TL performance, this paper proposes an innovative physical benchmark for PBSHM in the context of civil-engineering applications,

aimed at building a population of experimental model bridges where multiple bridge configurations are introduced by varying structural properties in terms of deck materials, number of piers and span lengths. An experimental campaign is carried out to build up a comprehensive dataset by simulating a variable number of damage scenarios and environmental conditions, since the mockup is located in an environmental chamber where a wide range of temperatures is reproduced. Such an experiment, representing a novelty in the literature, allows one to explore the applicability of PBSHM and underline the main advantage and possible drawbacks and specifically enables the use of DA, given that each bridge configuration is subjected to the same damage scenarios. To enable other researchers to test their own algorithms and address the various challenges associated with TL, this paper includes an open dataset containing all data related to the physical benchmark (https://doi.org/10.15131/shef.data.27732792). The results obtained after feature transformation show that different kinds of damages can be effectively identified across the population of bridges under environmental fluctuations, enabling discrimination between undamaged and damaged conditions. In some cases, the use of more labelled data from a different structure can also be useful to improve the identification of low-severity damage and, beyond that, the novel DA-based multi-source approach is demonstrated to yield a significant improvement on model's outcome.

However, additional insights should be made prior to DA. Specific studies on the influence of data-normalisation techniques (to remove environmental effects) on TL outcomes should be carried out, as well as on the relation between the bridge similarity and the type of damage (and thus the related features) to be transferred. Future work may also explore the possibility to investigate other TL strategies in order to address a wider portion of real-world scenarios in which DA assumptions can be broken. It could be interesting in the future to reproduce the transferability of health-state labels within real network-scale bridge monitoring, by taking into account the characteristics and the challenges of a population of post-tensioned continuous concrete bridges or continuous steel–concrete bridges. The validation on such structures will surely bring significant advances toward field applications and provide a real support to the traditional bridge maintenance. Finally, using archetypal numerical or physical models to generate extensive labelled data for various damage and degradation conditions offers a promising approach to identify critical conditions across multiple similar target bridges.

CRediT authorship contribution statement

Valentina Giglioni: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Jack Poole:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Robin Mills:** Supervision, Investigation, Conceptualization. **Ilaria Venanzi:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Filippo Ubertini:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Keith Worden:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

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

Acknowledgements

This work was supported by the Italian Ministry of University and Research (MUR) through the funded project of national interest "TIMING – Time evolution laws for IMproving the structural reliability evaluation of existING post-tensioned concrete deck bridges" (Protocol No. P20223Y947). The Authors would also like to acknowledge the UK Engineering and Physical Sciences Research Council (EPSRC) via grant EP/W005816/1.

Appendix. Summary of damage detection results via DA

See Table A.10.

Table A.10
Summary of damage detection results via DA for each pair of configurations.

d	B2-B1			B1-B2			B3-B1			B1-B3		
	TPr	FPr	F1	TPr	FPr	F1	TPr	FPr	F1	TPr	FPr	F1
M1	1	0.03	0.97	0	0.02	0.49	1	0.09	0.93	0.75	0	0.9
M2	0.8	0.03	0.9	0.83	0.04	0.78	0.9	0.03	0.94	0	0.05	0.42
M3	1	0	1	1	0.08	0.76	1	0	1	1	0	1
M4	1	0	1	0	0.05	0.86	1	0	1	1	0	1
M5	0	0	0.47	1	0.06	0.83	–	–	–	–	–	–
M6	0	0	0.47	1	0.06	0.75	–	–	–	–	–	–
M7	0.6	0	0.86	1	0.03	0.85	–	–	–	–	–	–
M8	0.2	0	0.64	1	0.06	0.75	–	–	–	–	–	–

(*continued on next page*)

Table A.10 (*continued*).

d	B4-B2			B2-B4			B4-B1			B1-B4		
	TPr	FPr	F1	TPr	FPr	F1	TPr	FPr	F1	TPr	FPr	F1
M1	0	0.007	0.49	0	0.005	0.49	0	0	0.44	0	0.04	0.47
M2	0.17	0	0.64	0	0.03	0.49	0	0	0.43	1	0.07	0.72
M3	1	0	1	0.9	0	0.97	1	0	1	1	0.06	0.8
M4	1	0.007	0.97	0.2	0.005	0.64	1	0	1	1	0.06	0.8
M5	1	0.01	0.95	0.5	0.009	0.78	0	0	0.46	1	0.01	0.92
M6	1	0	1	0.86	0.005	0.86	0	0	0.46	1	0.03	0.81
M7	1	0	1	0.75	0.005	0.87	1	0	1	1	0.03	0.76
M8	1	0	1	1	0	1	0.6	0	0.86	1	0.04	0.81
SB1	0	0	0.48	0.2	0.02	0.6	–	–	–	–	–	–
SB3	0.57	0	0.86	0.9	0.004	0.95	–	–	–	–	–	–
d	B4-B3			B3-B4			B3-B2			B2-B3		
	TPr	FPr	F1	TPr	FPr	F1	TPr	FPr	F1	TPr	FPr	F1
M1	0	0	0.41	1	0.1	0.71	1	0.1	0.67	0.75	0.05	0.86
M2	0.8	0	0.93	1	0.06	0.72	1	0.08	0.74	1	0	1
M3	0.88	0	0.95	1	0.03	0.86	1	0.02	0.91	1	0	1
M4	0.6	0	0.83	1	0.06	0.8	1	0.06	0.83	0.9	0	0.96

Data availability

As mentioned at the end of the manuscript, the data supporting the findings of this study will be openly available in a specific repository folder.

References

[1] Z. He, W. Li, H. Salehi, H. Zhang, H. Zhou, P. Jiao, Integrated structural health monitoring in bridge engineering, Automat. construct. 136 (2022) 104168.
[2] R.R. Rabi, M. Vailati, G. Monti, Effectiveness of vibration-based techniques for damage localization and lifetime prediction in structural health monitoring of bridges: A comprehensive review, Buildings 4 (2024) 1183.
[3] R. Zinno, S.S. Haghshenas, G. Guido, A. Vitale, Artificial intelligence and structural health monitoring of bridges: A review of the state-of-the-art, IEEE Access 10 (2022) 88058–88078.
[4] V.M. Di Mucci, A. Cardellicchio, S. Ruggieri, A. Nettis, V. Renò, G. Uva, Artificial intelligence in structural health management of existing bridges, Autom. Constr. 167 (2024).
[5] V. Giglioni, I. Venanzi, A.E. Baia, V. Poggioni, A. Milani, F. Ubertini, Deep autoencoders for unsupervised damage detection with application to the Z24 benchmark bridge, in: European Workshop on Structural Health Monitoring, Springer, 2022, pp. 1048–1057.
[6] A. Afshar, G. Nouri, S. Ghazvineh, S.H. Hosseini Lavassani, Machine-learning applications in structural response prediction: A review, Practice Period. Struct. Design Construct. 29 (3) (2024) 03124002.
[7] X. Zhao, S. Li, H. Su, L. Zhou, K.J. Loh, Image-based comprehensive maintenance and inspection method for bridges using deep learning, in: ASME 2018 Conference on Smart Materials, Adaptive Structures and Intelligent Systems, Vol. 2, SMASIS 2018, 2018, V002T05A017.
[8] X. Chen, J. Jia, J. Yang, Y. Bai, X. Du, A vibration-based 1DCNN-bilstm model for structural state recognition of RC beams, Mech. Syst. Signal Process. 203 (2023).
[9] M.A. Bud, I.D. Moldovan, M. Nedelcu, E. Figueiredo, Hybrid training of supervised machine learning algorithms for damage identification in bridges, in: European Workshop on Structural Health Monitoring, Springer, 2022, pp. 482–491.
[10] S. Sony, S. Gamage, A. Sadhu, J. Samarabandu, Vibration-based multiclass damage detection and localization using long short-term memory networks, in: Structures, Vol. 35, 2022, pp. 436–451.
[11] D. Xu, X. Xu, M.C. Forde, A. Caballero, Concrete and steel bridge structural health monitoring—Insight into choices for machine learning applications, Constr. Build. Mater. 402 (2023) 132596.
[12] K. Worden, L.A. Bull, P. Gardner, J. Gosliga, T.J. Rogers, E.J. Cross, E. Papatheou, W. Lin, N. Dervilis, A brief introduction to recent developments in population-based structural health monitoring, Front. Built Environ. 6 (2020) 146.
[13] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (10) (2009) 1345–1359.
[14] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, A comprehensive survey on transfer learning, Proc. IEEE 109 (1) (2020) 43–76.
[15] Q. Pan, Y. Bao, H. Li, Transfer learning-based data anomaly detection for structural health monitoring, Struct. Health Monit. 22 (5) (2023) 3077–3091.
[16] M.M. Azad, P. Kumar, H.S. Kim, Delamination detection in CFRP laminates using deep transfer learning with limited experimental data, J. Mater. Res. Technol. 29 (2024) 3024–3035.
[17] N. Bao, T. Zhang, R. Huang, S. Biswal, J. Su, Y. Wang, A deep transfer learning network for structural condition identification with limited real-world training data, Struct. Control Health Monit. 2023 (1) (2023) 8899806.
[18] S. Teng, X. Chen, G. Chen, L. Cheng, Structural damage detection based on transfer learning strategy using digital twins of bridges, Mech. Syst. Signal Process. 191 (2023) 110160.
[19] B. Zhang, X. Hong, Y. Liu, Multi-task deep transfer learning method for guided wave-based integrated health monitoring using piezoelectric transducers, IEEE Sens. J. 20 (23) (2020) 14391–14400.
[20] H.-P. Wan, Y.-Q. Ni, Bayesian multi-task learning methodology for reconstruction of structural health monitoring data, Struct. Health Monit. 18 (4) (2019) 1282–1309.
[21] L.A. Bull, D. Di Francesco, M. Dhada, O. Steinert, T. Lindgren, A.K. Parlikad, A.B. Duncan, M. Girolami, Hierarchical Bayesian modeling for knowledge transfer across engineering fleets via multitask learning, Comput.-Aided Civ. Infrastruct. Eng. 38 (7) (2023) 821–848.
[22] P. Gardner, L. Bull, J. Gosliga, N. Dervilis, K. Worden, Foundations of population-based SHM, part III: Heterogeneous populations–mapping and transfer, Mech. Syst. Signal Process. 149 (2021) 107142.

[23] M. Omori Yano, E. Figueiredo, S. da Silva, A. Cury, I. Moldovan, Transfer learning for structural health monitoring in bridges that underwent retrofitting, Buildings 13 (9) (2023) 2323.
[24] E. Figueiredo, M. Omori Yano, S. Da Silva, I. Moldovan, M. Adrian Bud, Transfer learning to enhance the damage detection performance in bridges when using numerical models, J. Bridge Eng. 28 (1) (2023) 04022134.
[25] P. Gardner, L.A. Bull, N. Dervilis, K. Worden, Domain-adapted Gaussian mixture models for population-based structural health monitoring, J. Civil Struct. Health Monitor. 12 (6) (2022) 1343–1353.
[26] J. Poole, P. Gardner, N. Dervilis, L. Bull, K. Worden, On statistic alignment for domain adaptation in structural health monitoring, Struct. Health Monit. 22 (3) (2023) 1581–1600.
[27] V. Giglioni, J. Poole, I. Venanzi, F. Ubertini, K. Worden, A domain adaptation approach to damage classification with an application to bridge monitoring, Mech. Syst. Signal Process. 209 (2024) 111135.
[28] Z.-D. Li, W.-Y. He, W.-X. Ren, Y.-L. Li, Y.-F. Li, H.-C. Cheng, Damage detection of bridges subjected to moving load based on domain-adversarial neural network considering measurement and model error, Eng. Struct. 293 (2023) 116601.
[29] L. Bull, P. Gardner, N. Dervilis, E. Papatheou, M. Haywood-Alexander, R. Mills, K. Worden, On the transfer of damage detectors between structures: An experimental case study, J. Sound Vib. 501 (2021) 116072.
[30] P. Gardner, L. Bull, J. Gosliga, J. Poole, N. Dervilis, K. Worden, A population-based SHM methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings, Mech. Syst. Signal Process. 172 (2022) 108918.
[31] M. Long, J. Wang, G. Ding, J. Sun, P.S. Yu, Transfer feature learning with joint distribution adaptation, Proc. IEEE Int. Conf. Computer Vision (2013) 2200–2207.
[32] O. Kramer, Dimensionality reduction with unsupervised nearest neighbors, Springer, 2013.
[33] E. García-Macías, F. Ubertini, MOVA/MOSS: two integrated software solutions for comprehensive structural health monitoring of structures, Mech. Syst. Signal Process. 143 (2020) 106830.
[34] J. Poole, P. Gardner, N. Dervilis, J. Mclean, T. Rogers, K. Worden, Towards physics-based metrics for transfer learning in dynamics, in: Society for Experimental Mechanics Annual Conference and Exposition, Springer, 2023, pp. 73–81.