13th International Conference on
Structural Health Monitoring of Intelligent Infrastructure
DOI: 10.3217/978-3-99161-057-1-136

A Transfer Learning approach for damage identification in operational viaducts

Eleonora Morleo [1, 0009-0002-5034-8833], Maria Pina Limongelli [1, 0000-0002-9353-5439], Andrea Piscini [2, 0000-0003-2786-1321], Edoardo Troielli [2, 0009-0007-8761-4361]

[1]Department of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy
[2]SINA S.p.A, Milan, Italy
email: eleonora.morleo@polimi.it, mariagiuseppina.limongelli@polimi.it, andrea.piscini@sina.it, edoardo.troielli@sina.it

ABSTRACT: A major limitation in data-driven Structural Health Monitoring is the scarcity of labeled data for training machine learning models. Transfer Learning addresses this by enabling knowledge sharing across similar structures, reducing datasets distribution shift. This study proposes a novel Transfer Learning framework for damage identification in operational viaducts with similar spans, using modal frequencies as damage-sensitive features. Domain Adaptation is performed via Normal Condition Alignment, to map source and target features in a shared latent space. A baseline normal condition is established on source features through a linear regression model. Gaussian Mixture Models are trained on source residuals, and used to detect anomalies in the target domain, based on residual distributions. A real viaduct for which long-term monitoring data are available is used as a case study. The structure comprises two homogeneous datasets collected on the deck of similar spans. Source data pertain to a deck with extensive measurements, whereas target data refer to a second deck with a reduced dataset, due to sensor malfunctions. Damage is simulated in the target dataset by reducing the measured frequencies. Validation using data from real damaged scenarios will enable future scaling of the proposed framework to operational conditions, providing a practical tool for data-driven SHM of viaducts, enabling damage detection in under-instrumented areas by leveraging data from other spans.

KEY WORDS: Domain adaptation; transfer learning; operational viaduct; SHM; anomaly detection; GMM; linear regression; temperature variations.

1 INTRODUCTION

The deterioration of infrastructures poses a considerable challenge to the safety and efficiency of Europe's roadway network, as evidenced by the significant number of bridges that are nearing the conclusion of their service life [1]. The implementation of permanent monitoring systems, designed to observe the evolution of structural behavior over time, is imperative to support effective management of the structure, planning of maintenance activities and timely interventions. However, the financial implications associated with the deployment of such systems across an entire bridge network constitute a major obstacle to the widespread implementation of continuous monitoring campaigns. Consequently, less costly alternatives are often considered, as point-in-time monitoring campaigns or the selective installation of sensors in critical sections or structural components.

To tackle the limitations imposed by cost-saving measures and limited investments, while still guaranteeing an adequate level of safety, a promising solution involves transferring knowledge acquired from heavily instrumented structures or components to others with limited data availability to support the assessment of their structural behavior. This can be achieved through Transfer Learning (TL), a method which has recently gained significant attention among researchers in the field of civil infrastructure monitoring, as a means to overcome the issue of scarcity of labelled data for operating structures [2], [3], [4]. The concept of leveraging data – and, by extension, knowledge – from multiple structures to inform inferences about a target structure was introduced for the first time within the Population-Based SHM framework [5], [6], [7], [8], [9]. Notably, Poole et al. [10] proposed a domain adaptation approach, referred to as Normal Condition Alignment (NCA), that aligns source and target datasets in a shared latent space, preserving the intrinsic meaning of damage-sensitive features, a capability not offered by other non-statistical approaches, such as DL. This statistical alignment technique has proven effective in several works concerning experimental and numerical datasets [11], [12], [13], in supervised settings with labelled data.

This paper investigates the application of NCA to an operational bridge equipped with a permanent monitoring system. Data from a well-instrumented span are leveraged to allow anomaly detection on a second span, where sensor failures have resulted in intermittent data acquisition.

The novelty of the proposed approach consists in the integration of a simple yet effective temperature compensation procedure following domain adaptation, to mitigate the confounding effect of temperature-induced variability on modal frequencies.

In the considered case study, no damage has been identified through field inspections, thereby the data from the source span are assumed to represent the undamaged state. For the target span, only the first year of monitoring data is considered representative of the undamaged configuration. Damage scenarios are synthetically introduced into the target dataset by increasingly reducing natural frequencies, to perform a sensitivity analysis of the proposed anomaly detection framework.

After domain adaptation is performed, a baseline normal condition for the source domain is established using a linear regression to model the relationship between temperature and natural frequencies. Residuals from this model are

DOI: 10.3217/978-3-99161-057-1-136

subsequently employed to train a Gaussian Mixture Model (GMM) that is then used for anomaly detection in the target domain. The proposed approach is compared to a baseline method, where the GMM is trained directly on domain-adapted features of the source domain, and tested on domain-adapted features of the target. The results, evaluated in terms of F1-score, indicate that training the GMM on residuals improves performance, particularly when a moderate temperature-frequency correlation is present.

This study demonstrates the effectiveness of a simple domain adaptation technique and regression framework in enhancing novelty detection for bridge monitoring. The proposed framework offers a practical and cost-effective solution for optimizing the use of available monitoring data for critical infrastructure integrity management.

The remainder of the paper is organized as follows: Section 2 illustrates the methodology, Section 3 describes the case study and the datasets, Section 4 discusses the results, and the final section provides conclusion remarks.

2 METHODOLOGY

2.1 Domain Adaptation via NCA

Traditional machine learning algorithms rely on the assumption that training and testing data originate from the same underlying distribution [14]. When transferring knowledge from a source domain to a target domain, this assumption not necessarily holds, due to possible differences in the statistical distributions of the two domains. These differences may lead to degraded performance and increased generalization error. Domain Adaptation addresses this limitation mitigating the distribution shifts between source and target domains, aligning feature distribution into a shared latent space.

In this work, source and target domains are aligned into a common feature space exploiting NCA. This statistical alignment technique incorporates prior knowledge about the health state of the structure: alignment is achieved by matching both the mean and standard deviation of the normal-condition data distributions to effectively reduce domain discrepancy between the source and the target datasets.

Let $D_S = \{\boldsymbol{x}_{s,i}, y_{s,i}\}_{i=1}^{n_s}$ be the source domain, with $\boldsymbol{x}_{s,i}$ the source feature vector, n_S the number of observations and $y_{s,i}$ the labels associated with each instance of the feature vector. Let also $D_T = \{\boldsymbol{x}_{t,l}, y_{t,l}\}_{l=1}^{n_t}$ be the target domain, with $\boldsymbol{x}_{t,l}$ the target feature vector, n_t the number of observations and $y_{t,l}$ the labels associated with each instance of the feature vector in the target domain. NCA is developed in two steps. First, the source domain is standardized:

$$\boldsymbol{z}_{s,i} = \frac{\boldsymbol{x}_{s,i} - \boldsymbol{\mu}_s}{\boldsymbol{\sigma}_s} \tag{1}$$

being $\boldsymbol{\mu}_s, \boldsymbol{\sigma}_s$ respectively the mean and the standard deviation of the source domain dataset. Then, the subset of the source domain consisting only of healthy instances $\boldsymbol{x}_{s,n,i}$ is considered, and its statistics are computed $\boldsymbol{\mu}_{s,n}, \boldsymbol{\sigma}_{s,n}$.

Assuming that the target domain includes data associated with the undamaged configuration, reasonably identified as those collected from the start of the monitoring activity up to one year later, the statistics of this subset are computed and the following transformation is applied:

$$\boldsymbol{z}_{t,l} = \frac{\boldsymbol{x}_{t,l} - \boldsymbol{\mu}_{t,n}}{\boldsymbol{\sigma}_{t,n}} \boldsymbol{\sigma}_{s,n} + \boldsymbol{\mu}_{s,n} \tag{2}$$

being $\boldsymbol{\mu}_{t,n}, \boldsymbol{\sigma}_{t,n}$ respectively the mean and the standard deviation of the target undamaged subset.

In this paper we assume that the entire source dataset is referred to the undamaged configuration of the structure, hence resulting in $\boldsymbol{\sigma}_{s,n} = \mathbf{1}$ and $\boldsymbol{\mu}_{s,n} = \mathbf{0}$.

2.2 Linear regression

Natural frequencies are sensitive to the influence of environmental and operational fluctuations, being temperature the most significant source of variability. To filter out the effect of temperature on the modal properties of the structure, regression models have been widely proven effective, being able to characterize the relationship between temperature and eigenfrequencies [15], [16].

Considering the frequencies time history depicted in Figure 2, a linear dependency of the modal parameters on temperature can be appreciated. Hence, following the alignment of source and target data in the latent feature space, we train a linear regression model on the source features $\boldsymbol{z}_{s,i}$ to learn the expected, temperature-dependent behavior of the structure. By doing so, the regression model establishes a baseline that characterizes the normal behavior of the source span. For each mode, the linear regression model reads as follows:

$$\hat{\boldsymbol{z}}_{s,i} = \beta_0 + \beta_1 \boldsymbol{T}_i \tag{3}$$

being $\hat{\boldsymbol{z}}_{s,i}$ the predicted source feature in the latent space, $\boldsymbol{T}_i$ the temperature corresponding to the *i-th* source feature $\boldsymbol{z}_{s,i}$ in the latent space, and β_0, β_1 the regression coefficients.

We then use this model to predict the features for both the source and the target domain, using the temperature values pertaining to each of them. The residuals, defined as the differences between the observed and predicted features values, are subsequently analyzed to detect potential anomalies or deviations from the expected structural behavior.

2.3 Gaussian Mixture Models for anomaly detection

Gaussian Mixture Models (GMMs) are probabilistic models employed in unsupervised machine learning for identifying clusters in data. They assume that data are generated from a mixture of multiple Gaussian distributions, each with unknown parameters. This assumption makes them suitable for modelling real-world monitoring data, which often exhibit multimodal distributions characterized by overlapping Gaussian components, due to changes in the data induced by factors such as temperature fluctuations and potential structural damage. Formally, the probability density function of the GMM is defined as:

$$p(\boldsymbol{x}) = \Sigma_{k=1}^{N} \pi_k N(\boldsymbol{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{4}$$

where π_k is the mixing coefficient (the weight) for the *k-th* Gaussian component, satisfying $0 \leq \pi_k \leq 1$ and $\Sigma_{k=1}^{N} \pi_k = 1$, and $N(\boldsymbol{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ is the *k*-th Gaussian distribution of the mixture, having $\boldsymbol{\mu}_k$ mean vector and $\boldsymbol{\Sigma}_k$ covariance matrix. The parameters of the GMM are estimated via the Expectation Maximization algorithm, which iteratively optimizes them by maximizing the likelihood of the observed data [14], [17].

GMMs are soft clustering algorithms, as they assign samples to a cluster based on the likelihood that each data point belongs

DOI: 10.3217/978-3-99161-057-1-136

to that cluster in particular. In this paper we use the GMMs for anomaly detection. Specifically, for each mode, we train a GMM on the source domain residuals. Once the model is fitted, the log-probability density function (log-PDF) of each instance is evaluated and a detection threshold is established based on a low quantile of the log-PDF distribution, to distinguish normal from anomalous data points. The trained GMM is then applied to the target domain residuals and the log-PDF is computed and compared to the previously defined threshold. Instances falling below this threshold are identified as anomalies, indicating a deviation from the normal behavior modeled using source data.

Last, we compare the results of the proposed anomaly detection framework, combining linear regression and GMM, to those obtained using GMM alone. This is to assess the effectiveness of the integrated approach versus a direct application of GMM. More specifically, a separate GMM is trained, for each mode, using a dataset comprising the frequency and temperature vectors associated with the source domain in the latent space. Each GMM employs full covariance matrices to capture the correlations between frequency and temperature, and the optimal number of components for each model is selected based on the Akaike Information Criterion (AIC), to balance model complexity and goodness of fit. Then, for each mode, each GMM is tested against the target dataset in the latent space, including once again the frequency and temperature vectors associated with the selected mode.

2.4 Performance evaluation

To assess the performance of the proposed anomaly detection framework, the number of true positives (TP), true negatives (TN), false positives (FP), false negatives (FN) is computed. These quantities enable the evaluation of precision and recall metrics, which in turn are used to compute the F1 score. Precision measures the proportion of detected anomalies among all instances (TP / (TP + FP)), while recall measures the fraction of correctly detected anomalies (TP / (TP + FN)) [14]. Then, the harmonic mean of the two metrics is computed, resulting in the F1-score:

$$\mathrm{F1-score} = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (5)$$

The F1-score is a commonly used metric to assess the performance of classification algorithms in presence of unbalanced dataset, a common issue in SHM datasets, where the number of healthy instances often exceeds the number of damaged ones by far.

3 CASE STUDY

3.1 The viaduct

The structure under consideration in this study is an operational viaduct located in northern Italy. Constructed in 1968, the viaduct comprises 41 spans, extending for a total length of 1673.5 meters. The deck features continuous pre-stressed Gerber beams, each resting on two cast-in-place piers, and suspended Gerber spans (Figure 1). Of particular note is the monitoring of eleven spans that intersect with a river bed. In this section, the piers are spaced 61.5 meters apart, with a total length of 676.5 meters.

The instrumentation installed on the eleven monitored spans comprises a total of 219 biaxial MEMS inclinometers and 100 triaxial MEMS accelerometers, installed in 2020 to monitor the dynamic characteristics of the bridge and trigger alarms when fixed thresholds are exceeded. The accelerometers are not distributed uniformly across each span; rather, they are positioned on only half of each span.

The present study focuses on accelerometer data collected from May 2022 to December 2024. The structure is found to be in normal condition and is currently in service. Consequently, no damage configuration labels are available, as there is no evidence of damage to the structure. Temperature information is available for each of the set of eigenfrequencies obtained from the acceleration time histories.

3.2 Description of the dataset

The objective of this study is to leverage the large volume of monitoring data available for a single span of the viaduct (the source) to perform anomaly detection on another similar span (the target), with reduced data availability due to sensors malfunction. Two continuous spans are considered. Both the source and target dataset consist of time series observations that include the first three eigenfrequencies and the corresponding temperatures measured at each time step. Specifically, the source dataset comprises 4346 observations organized in a 4346 × 4 matrix, while the target dataset includes 2143 observations structured in an analogous 2143 × 4 matrix.

As previously stated, according to available inspection records, there is not any evidence of damage affecting the viaduct so far. Hence, both the source and target dataset considered contain observations which are reflective of the normal condition of the relative spans. For the source dataset, this condition is treated as factual throughout the study. Instead, we assume that only the observations acquired during the first year of monitoring of the target span are associated with its undamaged configuration. Additionally, we simulate the

Figure 1. Part of the longitudinal section of the viaduct. Span 34 is the source one. In magenta the accelerometers installed.

DOI: 10.3217/978-3-99161-057-1-136

occurrence of damage in the target by reducing the frequency values acquired from August 2024 by 2% and 5%. This enables us to conduct a sensitivity analysis on the performance of the anomaly detection procedure implemented.

The study is conducted in an unsupervised setting. However, to assess the performance of the framework in detecting anomalies in a real scenario, labels are assigned to the target dataset to mark the instances of reduced frequencies simulating damage, making it possible to compute the F1-score.

Figure 2 illustrates the first three natural frequencies for both source and target spans. The fundamental frequency *f1* is approximately equal to 1.5 Hz for both the spans, and exhibits a similar trend for both domains. The second natural frequency *f2* oscillates around 2.6 Hz. In this case, while there is a marked dependency of the source *f2* on temperature, the corresponding target frequency displays a smoother trend. Finally, the third natural frequency of the two spans differs significantly, with *f3* of the source domain having a mean value of 4.4 Hz, whereas *f3* of the target dataset is lower, with a mean value of 3.6 Hz. Furthermore, a strong sensitivity to daily temperature fluctuations can be appreciated in the third eigenfrequency of the source span, with respect to the target one. Figure 3 depicts the temperature measurements recorded by the accelerometers installed on both spans. Figure 4 illustrates the time histories of *f1*, *f2* and *f3* of the target domain, highlighting the artificially reduced frequencies starting from August 2024 onward to simulate damage.

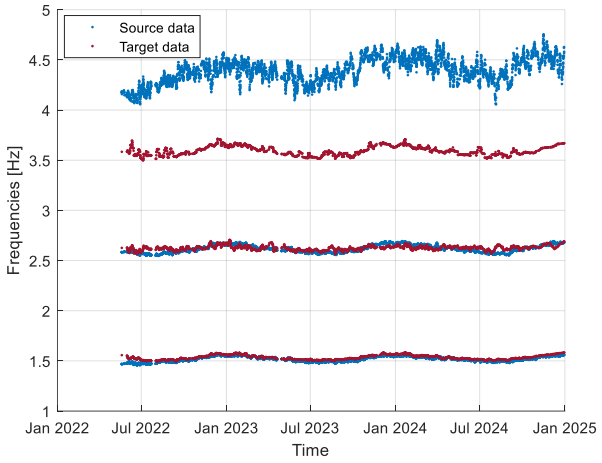


Figure 2. First three natural frequencies for source (in blue) and target dataset (in red). Frequencies f1 and f2 range around 1.5 and 2.6 Hz respectively, while f3 is around 3.6 Hz for the target domain and higher, around 4.3Hz, for the source domain.

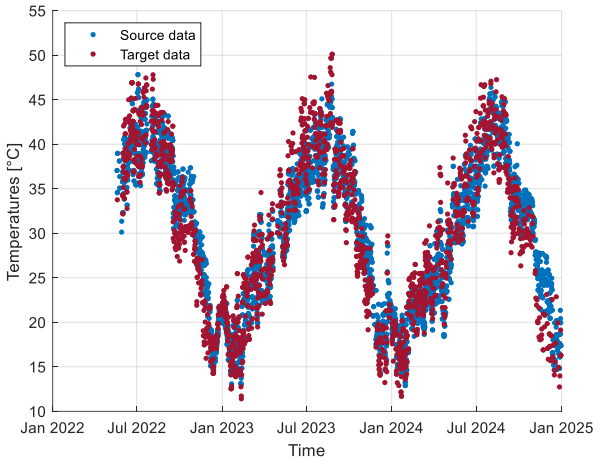
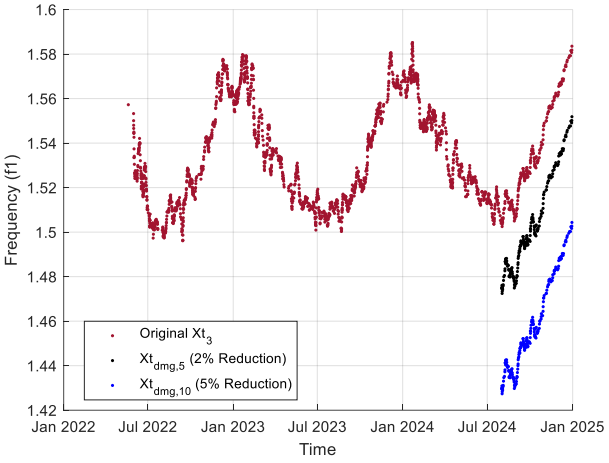
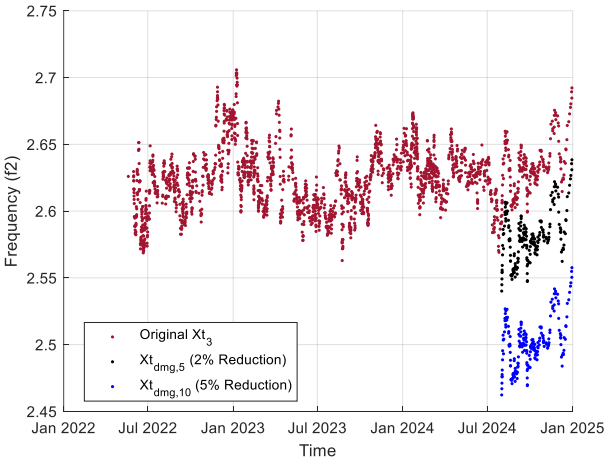


Figure 3. Source and target dataset temperature records.

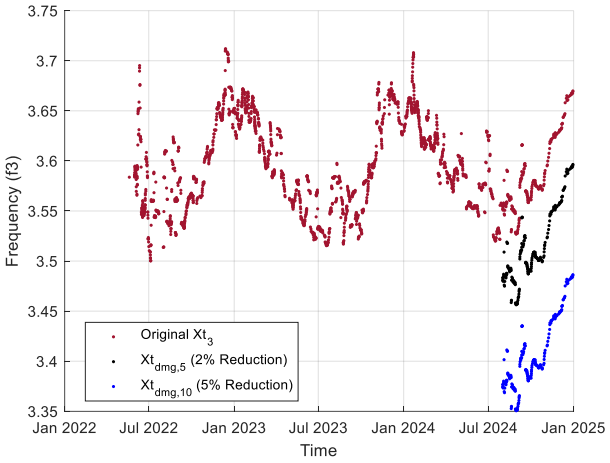


a)



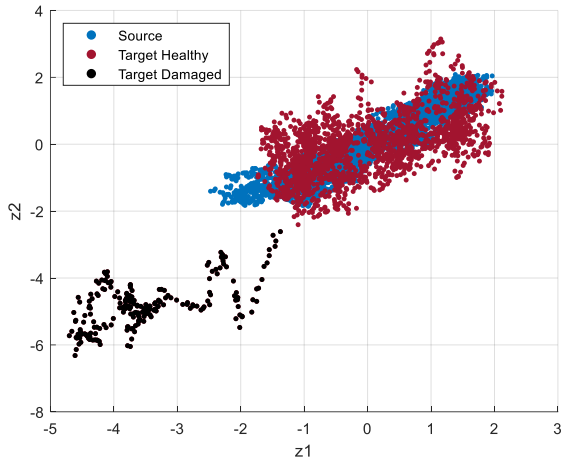
b)

DOI: 10.3217/978-3-99161-057-1-136



c)

Figure 4. Original and synthetic data for the target domain. Frequencies of the target domain are reduced by 2% and 5% starting from August 2024 to simulate damage. Subfigures a), b), c) refer respectively to the first, second, and third natural frequencies.



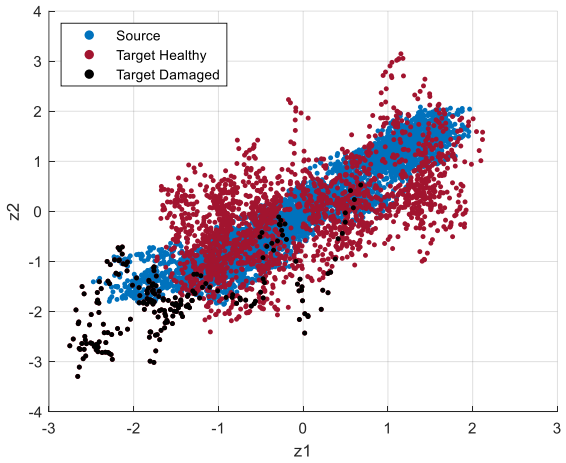
b)

Figure 5. Source and target features (z1 and z2) after domain adaptation. Figure 5a depicts the scenario in which damage is simulated through a 2% reduction in natural frequencies, whereas Figure 5b corresponds the case of a 5% frequency reduction.

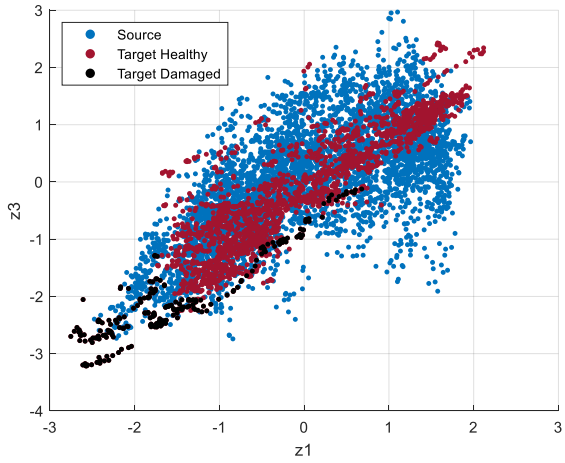
4 RESULTS

Normal Condition Alignment has been applied to the source and target datasets, and the resulting features in the latent space are depicted in Figure 5 and Figure 6. Figure 5a and Figure 6a are relative to the case where damage is simulated as a 2% reduction in natural frequencies, while Figure 5b and Figure 6b refer to a 5% frequencies reduction.

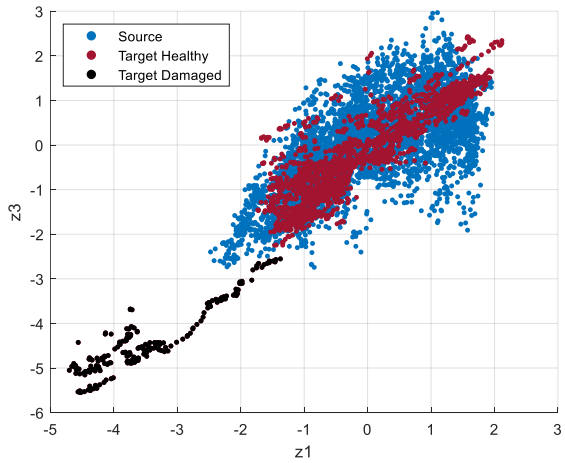
As a result of the application of NCA, when a 5% frequency reduction is introduced to simulate damage (Figure 5b and Figure 6b), we see that the normal instances of the target domain overlap the healthy features of the source. In contrast, the simulated damage instances tend to form distinct and distant clusters in the latent space. On the other hand, when damage is simulated through only a 2% reduction in frequencies, the corresponding damage instances of the target tend to cluster closely with, and even overlap, the healthy source and target features (Figure 5a and Figure 6a).



a)



a)



b)

Figure 6. Source and target features (z1 and z3) after domain adaptation. Figure 6a depicts the scenario in which damage is simulated through a 2% reduction in natural frequencies,

DOI: 10.3217/978-3-99161-057-1-136

whereas Figure 6b corresponds the case of a 5% frequency reduction.

In the following, the results associated with the 5% reduction in frequencies will be first illustrated, while those associated with the 2% frequency reduction for damage simulation will be discussed in the remainder of this section.

Separate linear regression models have been trained for each of the three source features (Figure 7, Figure 8, Figure 9) to capture their relationship with the corresponding temperature values. Each trained model has then been utilized to make predictions on both source data and target data, and the residuals between the actual feature values and the predicted ones have been computed.

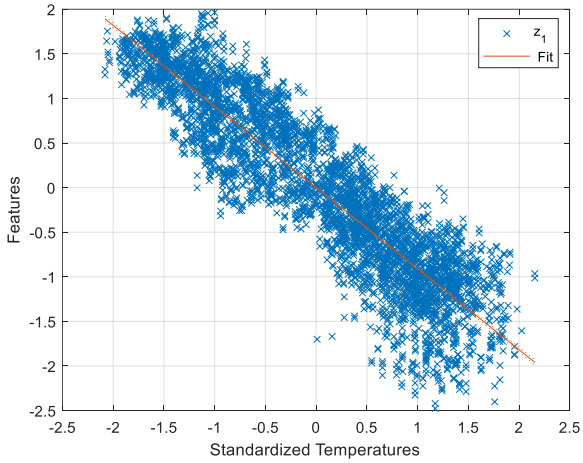


Figure 7. Linear regression of feature *z1* of the source domain vs temperature.

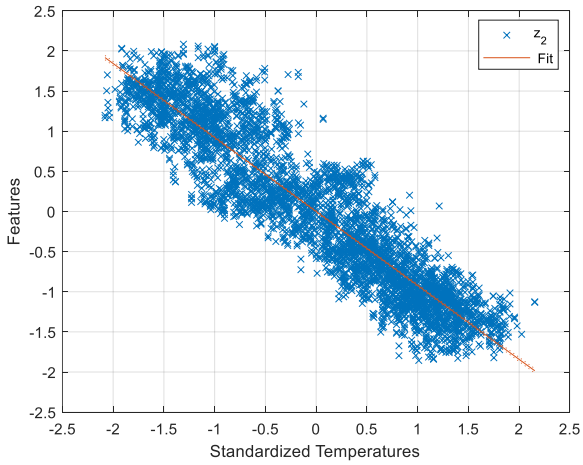


Figure 8. Linear regression of feature z2 of the source domain vs temperature.

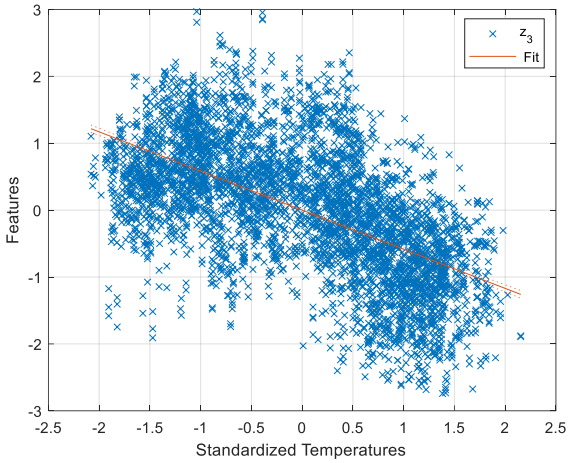


Figure 9. Linear regression of feature *z3* of the source domain vs temperature.

A Gaussian mixture model has been independently trained on the source residuals for each mode. Following the training, the log-PDF of each GMM was evaluated, and a detection threshold was established at the 5th quantile for modes 1 and 3, and at the 1st quantile for mode 2. Subsequently, the trained GMMs have been applied to the target residuals for the corresponding modes, enabling anomaly detection by comparing the log-PDF values for the target against the predefined thresholds. The results are illustrated in Figure 10, Figure 11, and Figure 12, and the performance metrics are listed in Table 1.

Table 1. Anomaly detection results using GMMs trained on source residuals, evaluated on the target dataset with synthetic damage simulated by a 5% reduction in natural frequencies.

	TP	FP	FN	TN	Precision	Recall	F1 score
f1	219	95	0	1829	0.70	1.00	0.82
f2	219	475	0	1449	0.32	1.00	0.48
f3	219	14	0	1919	0.94	1.00	0.97

As we can see from the table, the GMMs tested on target residuals are in all cases able to detect the true damage instances, while detecting few false positives for *z1* and *z3*. In the case of *z2*, the number of false positives is considerable, and this is probably due to the fact that frequency *f2* of the target does not exhibit the same strong temperature dependence that is inherently present in the second natural frequency of the source domain.

Comparing these results with those obtained by training, for each mode, a GMM directly on the source features and temperatures (Table 3) in the latent space, a slight improvement in the F1-score can be appreciated when residual analysis is applied. As a matter of fact, from Table 2 we can see that, even though also in this case all the true anomalies are correctly identified, the baseline GMM produces a slightly higher number of false positives with respect to the residual-based approach. This demonstrates that the proposed framework can enhance the anomaly detection process by reducing the false alarms that would otherwise arise from applying the GMM to latent variables directly.

DOI: 10.3217/978-3-99161-057-1-136

Table 2. Anomaly detection results using GMMs trained on source features directly, evaluated on the target dataset with synthetic damage simulated by a 5% reduction in natural frequencies.

	TP	FP	FN	TN	Precision	Recall	F1 score
f1	219	126	0	1798	0.63	1	0.78
f2	219	520	0	1404	0.30	1	0.46
f3	219	57	0	1867	0.79	1	0.88

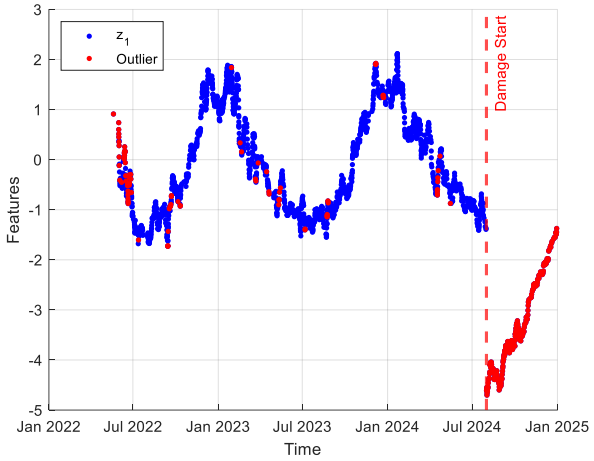


Figure 10. Target anomalies detected by GMM trained with source features residuals r1. The threshold is set to the 0.05 quantile of the log-PDF.

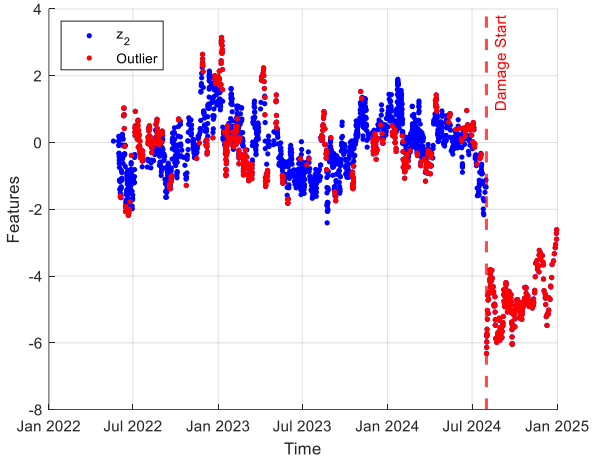


Figure 11. Target anomalies detected by GMM trained with source features residuals *r2*. The threshold is set to the 0.01 quantile of the log-PDF.

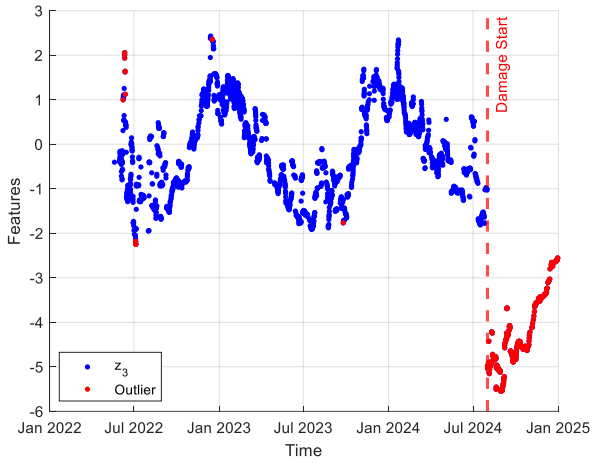


Figure 12. Target anomalies detected by GMM trained with source features residuals *r3*. The threshold is set to the 0.05 quantile of the log-PDF.

Table 3. F1 scores comparison on the target dataset with synthetic damage data, reduction of 5%.

5% reduction	GMM trained on source features	GMM trained on source residuals	F1-score improvement
f1	0.78	0.82	5.13 %
f2	0.46	0.48	4.35 %
f3	0.88	0.97	10.23 %

Finally, Table 4 illustrates the results for the scenario where damage is simulated via a reduction of the actual target frequency values by 2%. In this case, we can see that the performances of the GMMs trained on source residuals are slightly worse than in the case with the 5% frequency reduction, standing the same detection thresholds. Nevertheless, the F1-scores computed for the residual-based anomaly detection approach still exhibit higher values than those achieved by GMMs trained on source latent features (Table 5), registering an increase in the F1-score value up to 10.5% for the third natural frequency (Table 6). This outcome highlights that the proposed framework is effective even when the damage is subtle, helping to mitigate the number of false positives that can occur in such cases.

Table 4. Anomaly detection results using GMMs trained on source residuals, evaluated on the target dataset with synthetic damage simulated by a 2% reduction in natural frequencies.

	TP	FP	FN	TN	Precision	Recall	F1 score
f1	202	94	17	1830	0.68	0.92	0.78
f2	178	472	41	1452	0.27	0.81	0.41
f3	168	14	51	1910	0.92	0.77	0.84

DOI: 10.3217/978-3-99161-057-1-136

Table 5. Anomaly detection results using GMMs trained on source features directly, evaluated on the target dataset with synthetic damage simulated by a 2% reduction in natural frequencies.

	TP	FP	FN	TN	Precision	Recall	F1 score
f1	201	126	18	1798	0.61	0.92	0.74
f2	180	543	39	1381	0.25	0.82	0.38
f3	171	62	48	1862	0.73	0.78	0.76

Table 6. F1 scores comparison on target dataset with synthetic damage data, reduction 2%.

2% reduction	GMM trained on source features	GMM trained on source residuals	F1-score improvement
f1	0.74	0.78	5.41 %
f2	0.38	0.41	7.89 %
f3	0.76	0.84	10.53 %

5 CONCLUSIONS

In this paper, we have proposed a novel domain adaptation framework aimed at enhancing unsupervised anomaly detection in operational viaduct components. We have utilized the Normal Condition Alignment algorithm to leverage the extensive dataset available for a source span, to support anomaly detection in a target span with limited observations due to sensor malfunction. The novelty of this work lies in the use of a linear regression model to capture the dependency between frequency and temperature in the source domain after domain adaptation. This enables the establishment of a baseline normal condition for the source span in the latent space, which can then be "transferred" to the target span. Residuals from the source linear regression model have been used to train a Gaussian Mixture Model for each mode, enabling the detection of deviations from the expected normal behavior in the target domain when the trained models are tested against the target residuals.

We have compared the results of our residual-based framework against those obtained by training and testing the GMMs directly on domain-adapted features. The outcomes have highlighted that our method reduces the number of false positives, especially for modes where the natural frequencies are strongly influenced by temperature fluctuations, and even in the case in which the presence of damage is subtle, such as when frequencies from the normal condition are reduced by 2%. This is particularly relevant considering that even slight shifts in natural frequencies may correspond to severe structural damage.

Furthermore, from Figure 5a and Figure 6a we have seen that in the case of a 2% frequency reduction, the damaged instances of the target domain tend to overlap the healthy clusters of both source and target domains in the latent space. Relying solely on source features to train the GMM, without accounting for temperature in the establishment of a baseline normal condition, would have led to even poorer performances compared to using both features and temperatures. In fact, overlapping damaged instances would likely be misclassified as healthy, thus increasing the number of false negatives.

In conclusion, the findings of this study underscore the significance of explicitly incorporating environmental variability into the knowledge transfer process. Future research will focus on examining the applicability of the proposed framework in contexts involving dissimilar source and target spans, as well as datasets with substantial data gaps. Additionally, validating the framework with data from actual damaged scenarios will pave the way for its future application to operational conditions. This will support the development of a practical, data-driven SHM tool for viaducts, capable of detecting damage even in under-instrumented spans by leveraging information from other ones.

ACKNOWLEDGMENTS

The authors acknowledge SINA S.p.A. for sharing monitoring data and all relevant documentation related to the case study. Further, the authors thank Eray Temur for the development of the automated Operational Modal Analysis procedure used for extracting natural frequencies from the acceleration time histories.

REFERENCES

[1] Gkoumas, K. *et al.*, "Research and innovation in bridge maintenance, inspection and monitoring - A European perspective based on the Transport Research and Innovation Monitoring and Information System (TRIMIS)," vol. EUR 29650 EN, 2019, doi: 10.2760/719505.

[2] M. Omori Yano, E. Figueiredo, S. Da Silva, and A. Cury, "Foundations and applicability of transfer learning for structural health monitoring of bridges," *Mech. Syst. Signal Process.*, vol. 204, p. 110766, Dec. 2023, doi: 10.1016/j.ymssp.2023.110766.

[3] L. Souza, M. O. Yano, S. Da Silva, and E. Figueiredo, "A Comprehensive Study on Unsupervised Transfer Learning for Structural Health Monitoring of Bridges Using Joint Distribution Adaptation," *Infrastructures*, vol. 9, no. 8, p. 131, Aug. 2024, doi: 10.3390/infrastructures9080131.

[4] E. Figueiredo, M. Omori Yano, S. Da Silva, I. Moldovan, and M. Adrian Bud, "Transfer Learning to Enhance the Damage Detection Performance in Bridges When Using Numerical Models," *J. Bridge Eng.*, vol. 28, no. 1, p. 04022134, Jan. 2023, doi: 10.1061/(ASCE)BE.1943-5592.0001979.

[5] L. A. Bull *et al.*, "Foundations of population-based SHM, Part I: Homogeneous populations and forms," *Mech. Syst. Signal Process.*, vol. 148, p. 107141, Feb. 2021, doi: 10.1016/j.ymssp.2020.107141.

[6] J. Gosliga, D. Hester, K. Worden, and A. Bunce, "On Population-based structural health monitoring for bridges," *Mech. Syst. Signal Process.*, vol. 173, p. 108919, Jul. 2022, doi: 10.1016/j.ymssp.2022.108919.

[7] P. Gardner, L. A. Bull, J. Gosliga, N. Dervilis, and K. Worden, "Foundations of population-based SHM, Part III: Heterogeneous populations – Mapping and transfer," *Mech. Syst. Signal Process.*, vol. 149, p. 107142, Feb. 2021, doi: 10.1016/j.ymssp.2020.107142.

[8] J. Gosliga, P. A. Gardner, L. A. Bull, N. Dervilis, and K. Worden, "Foundations of Population-based SHM, Part II: Heterogeneous populations – Graphs, networks, and communities," *Mech. Syst. Signal Process.*, vol. 148, p. 107144, Feb. 2021, doi: 10.1016/j.ymssp.2020.107144.

[9] G. Tsialiamanis, C. Mylonas, E. Chatzi, N. Dervilis, D. J. Wagg, and K. Worden, "Foundations of population-based SHM, Part IV: The geometry of spaces of structures and their feature spaces," *Mech. Syst. Signal Process.*, vol. 157, p. 107692, Aug. 2021, doi: 10.1016/j.ymssp.2021.107692.

[10] J. Poole, P. Gardner, N. Dervilis, L. Bull, and K. Worden, "On statistic alignment for domain adaptation in structural health monitoring," *Struct. Health Monit.*, vol. 22, no. 3, pp. 1581–1600, May 2023, doi: 10.1177/14759217221110441.

[11] V. Giglioni, J. Poole, R. Mills, I. Venanzi, F. Ubertini, and K. Worden, "Transfer learning in bridge monitoring: Laboratory study on domain adaptation for population-based SHM of multispan continuous girder

DOI: 10.3217/978-3-99161-057-1-136

bridges," *Mech. Syst. Signal Process.*, vol. 224, p. 112151, Feb. 2025, doi: 10.1016/j.ymssp.2024.112151.
[12] V. Giglioni, I. Venanzi, and F. Ubertini, "Supervised machine learning techniques for predicting multiple damage classes in bridges," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2023*, Z. Su, M. P. Limongelli, and B. Glisic, Eds., Long Beach, United States: SPIE, Apr. 2023, p. 48. doi: 10.1117/12.2664359.
[13] V. Giglioni, J. Poole, I. Venanzi, F. Ubertini, and K. Worden, "A domain adaptation approach to damage classification with an application to bridge monitoring," *Mech. Syst. Signal Process.*, vol. 209, p. 111135, Mar. 2024, doi: 10.1016/j.ymssp.2024.111135.
[14] K. P. Murphy, *Machine Learning - A Probabilistic Perspective*. in Adaptive Computation and Machine Learning. Cambridge: MIT Press, 2014.
[15] B. Peeters, J. Maeck, and G. D. Roeck, "Vibration-based damage detection in civil engineering: excitation sources and temperature effects," *Smart Mater. Struct.*, vol. 10, no. 3, pp. 518–527, Jun. 2001, doi: 10.1088/0964-1726/10/3/314.
[16] F. Ponsi, G. E. Varzaneh, E. Bassoli, B. Briseghella, C. Mazzotti, and L. Vincenzi, "Temperature effect on the modal frequencies of a steel railway bridge," *Procedia Struct. Integr.*, vol. 62, pp. 1051–1060, 2024, doi: 10.1016/j.prostr.2024.09.140.
[17] C. M. Bishop, *Pattern recognition and machine learning*. in Information science and statistics. New York: Springer, 2006.