Multi-class Damage Detection on Experimental Frames through Transfer Component Analysis

e-Journal of Nondestructive Testing - ISSN 1435-4934 - www.ndt.net

EWSHM 2024
11th European Workshop on Structural Health Monitoring

Multi-class Damage Detection on Experimental Frames through Transfer Component Analysis

Alessio CROCETTI [1,2], Linda SCUSSOLINI [1], Gaetano MIRAGLIA [1,2], Rosario CERAVOLO [1,2]

[1]Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Turin, Italy; alessio.crocetti@polito.it, linda.scussolini@polito.it, gaetano.miraglia@polito.it, rosario.ceravolo@polito.it.

[2]Responsible Risk Resilience interdepartmental Centre (R3C), Politecnico di Torino, Turin, Italy; alessio.crocetti@polito.it, gaetano.miraglia@polito.it, rosario.ceravolo@polito.it.

Abstract. Machine learning techniques have been revealed to be useful in damage identification applications, exploiting, for example, algorithms able to detect anomalies in the monitored data. However, these datasets are not always enough to train a damage classifier reliably. This happens because the lack, or the low number of labelled data, produces outcomes prone to overfitting and bias (due to the low attainable statistical significance of the dataset). To avoid this problem, Transfer Learning technique can be exploited to make up for the lack of data available on a structure (target) using data recorded on another structural system (source), more or less similar, which is rich in data referring to the same structural behavior to be classified (e.g., damage detection in columns due to crack propagation).

In this work, a specific sub-application of Transfer Learning, named Transfer Component Analysis technique, is exploited on a benchmark system (numerically or experimentally) represented by a three-story aluminum scaled structure, subjected to increasing damage and mass variation over the three different floors. Special emphasis will be given to how accuracy is affected by data distribution, and in more detail, the authors will show how the increase in accuracy is related to the type of damage to be classified.

Keywords: Damage identification; Machine Learning; Domain adaptation; Transfer Learning; Multi-class damage detection.

https://doi.org/10.58286/29791

Introduction

Detecting structural damage is a crucial aspect of Structural Health Monitoring (SHM) across civil, mechanical, and aerospace engineering. Modal parameters such as natural frequencies provide vital insights into structural characteristics like stiffness and mass. Continuous monitoring of these parameters aims to detect deviations indicating structural anomalies. The increasingly manageable availability of monitoring data is leading to a notable need to rely on efficient computationally algorithms that are able to process data and provide a result that is easily interpretable by SHM technicians. To address this, Machine Learning (ML) techniques are utilized to analyze the data without requiring a numerical model of the structure. In this regard, structural monitoring can be classified into two broad approaches: *model-based* and *data-driven*.

Model-based SHM relies on the use of mathematical models, capable to simulate the physical behavior in presence of specific issues (e.g., structural, thermal, etc.). To this aim, often Finite Element (FE) models [1], that discretize the continuum and interpolate the results with specific polynomial functions, are employed. The advantage of this approach lies in the variety of correlated data, providing a causal relationship between the quantity of input and output. However, these models are often difficult to computationally manage and calibrate with respect to experimental observations [2], [3].

On the other hand, data-driven SHM employes regression and classification models (such as statistical and probabilistic ML models) between input and output data to recognize a pattern (Pattern Recognition – PR, [4]). They allow generalization, but the causal relationship between input and output is not guaranteed. Moreover, the interpretation of the results is often limited to specific observed problems (i.e., limited context of engineering interest). They have the great advantage of being extremely light from a computational point of view, and completely scalable to any type of problem. Structural health assessment is usually based on symptoms.

One significant challenge in data-driven approaches is the lack of data representing damaged states, which hampers pattern recognition processes. Transfer Learning (TL) addresses this issue by leveraging data from a more accessible system (source) to initialize algorithms for systems with limited data (target) but similar properties [5], [6].

Research significance

Most of the time, when performing SHM tasks on real structures, the lack of labelled data on the presence of damage and damage typology brings to a high uncertainty in providing instructions for maintenance and early intervention on the structure. On real structures is quite expensive (and often impossible) gather data on the damaged condition because it would mean breaking down the structure on field. This is even more true if the information is necessary for different types of damage, as the enforcement of one type of damage would influence the next. The present work prompts to solve this experimental issue by exploiting numerical models within a TL approach. In more detail, numerical and experimental datasets of an aluminum three-story frame are considered to perform a multi-class damage detection problem. Here, Transfer Component Analysis (TCA) [7], belonging to the domain adaptation methods, is used to exploit numerical source data with the aim of improving damage detection and classification on the experimental target data. In doing this, the authors propose the use of both natural frequencies and mode shapes information, the latter helping in providing spatial information on damage.

Section 1 deals with the methods, in particular referring to the TCA algorithm. Section 2 describes the application of the methods on a three-story aluminum structure: here the different damaged conditions together with the undamaged one are described. On Section

3 the TCA improvement on the accuracy of the multi-class damage detection problem for the target system are clearly depicted. Finally, general conclusions are reported on Section 4.

1. Methods

1.1 Transfer Component Analysis

Domain Adaptation is a sub-sector of TL methods that aims to transfer knowledge between two related systems by leveraging associated data domains [7], [8]. The term *domain* can be explained through two different components: the feature space of input $\mathcal{X}$ and the Marginal distribution $P(X)$ of a set of inputs $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}^T \in \mathcal{X}$, with $X \in \mathbb{R}^{n\times d}$, where n is the observation space and d the dimension of features. Within the framework of TL, two distinct domains are defined: a source domain $\mathcal{D}_\mathrm{S}$ and a target domain $\mathcal{D}_\mathrm{T}$. Each domain is associated with a task $\mathcal{T} = \{\mathcal{Y}; f(\cdot)\}$, where $\mathcal{Y}$ represents the label space and $f(\cdot)$ denotes the objective predictive function used for predicting the corresponding label. A common way to define this function is $P(\boldsymbol{y}|X)$, with $\boldsymbol{y} \in \mathcal{Y}$. In the domain adaption methods, the feature and label spaces are considered to be identical for both source and target domains, i.e., having $\mathcal{X}_\mathrm{S} = \mathcal{X}_\mathrm{T}$ and $\mathcal{Y}_\mathrm{S} = \mathcal{Y}_\mathrm{T}$. In the SHM field it means that both systems share the same diagnostic properties and the same structural conditions. However, in reality, the marginal distribution differs between the two domains, and not always in the conditional distribution. These differences imply that the diagnostic features are distributed differently ($P(X_S) \neq P(X_T)$), where $X_S \in \mathbb{R}^{n_S\times d}$, $X_T \in \mathbb{R}^{n_T\times d}$ with n_S and n_T being the source and target observations. Moreover, it also expresses the differences in the probabilities related to the occurrence of structural conditions (labels), $P(Y_S|X_S) \neq P(Y_T|X_T)$, meaning dissimilarities in features happening between the two domains. Training a classifier on the source domain and directly testing it on the target domain may induce in error the algorithm. To address this challenge, several methods have been proposed to reduce the distribution gap between source and target domains. These techniques typically involve the use of a nonlinear mapping function $\phi(\cdot)$ to align the distribution, resulting in $P(\phi(X_S)) \approx P(\phi(X_T))$ and $P(Y_S|\phi(X_S)) \approx P(Y_T|\phi(X_T))$. In this paper, Transfer Component Analysis (TCA) algorithm is employed to minimize the distance between data distributions.

TCA [7] operates under the assumption that $P(X_S) \neq P(X_T)$, but $P(Y_S|X_S) = P(Y_T|X_T)$, seeking to find a mapping function $\phi(\cdot)$ from the source space to a Reproducing Kernel Hilbert Space (RKHS) using Maximum Mean Discrepancy (MMD) as embedded criterion. This aims to minimize the distance between the marginal probabilities $P(\phi(X_S))$ and $P(\phi(X_T))$, while ensuring $P(Y_S|\phi(X_S)) \approx P(Y_T|\phi(X_T))$. The function ϕ can be determined as a feature map defined by a universal kernel. The MMD distance between the two data distributions is measured through the distance between the empirical means of the source and target domains:

$$Dist(X'_S, X'_T) = tr(KL) \tag{1}$$

where X' represent the transformed inputs, $K = k(X, X') \in \mathbb{R}^{(n_S+n_T)\times(n_S+n_T)}$ is associated with the Kernel matrix, $X = \{X_S, X_T\}^T$, and L is the MMD matrix, defined as $L(i,j) = 1/n_S^2$ if $\boldsymbol{x}_i, \boldsymbol{x}_j \in X_S$, else $L(i,j) = 1/n_T^2$ if $\boldsymbol{x}_i, \boldsymbol{x}_j \in X_T$, otherwise $L(i,j) = -(1/n_S n_T)$ [9]. By exploiting a weight matrix $W \in \mathbb{R}^{(n_S+n_T)\times m}$, with m the reduced dimensional space of the feature vector, through a kernel matrix decomposition the empirical kernel map becomes $\widetilde{K} = KWW^TK$ [10]. Substituting $\widetilde{K}$, the distance can be defined as:

$$Dist(X'_S, X'_T) = tr(W^TKLKW) \tag{2}$$

A regularization term is introduced in the distance minimization with the aim of control W, reformulating the kernel problem avoiding the trivial solution $W = 0$ introducing the constraints:

$$\min_{s.t.W^THKW=I} tr(W^TKLKW) + \mu\, tr(W^TW) \qquad 3$$

where μ is a regularization/trade off parameter, I is the identity matrix, $H = I - 1/(n_S + n_T)\mathbf{1}$ is a centering matrix, with $\mathbf{1}$ representing a matrix of ones, and W^THKW corresponding to the variance of the projected samples preserved in TCA. By writing the Lagrangian, the latter equation can be solved by optimizing the equivalent trace problem in the following:

$$\min_{s.t.W^THKW=I} tr(W^TKLKW) + \mu\, tr(W^TW) \qquad 4$$

W can be solved by computing the m principal eigenvectors of $(KLK + \mu I)^{-1}KHK$, where $m \leq n_S + n_T - 1$, describing the space of transformed features by means of $Z = KW \in \mathbb{R}^{(n_S+n_T)\times m}$.

To evaluate the performance of the algorithm in the classification process, an accuracy metric is defined based on the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad 5$$

In conclusion, a confusion matrix [11] is utilized to depict the comparison between the true class (ground truth) representing a particular damaged or undamaged condition and the class predicted by the algorithm, evaluating accuracy within each cell.

2. Application

2.1 Experimental and numerical three-story aluminum frame

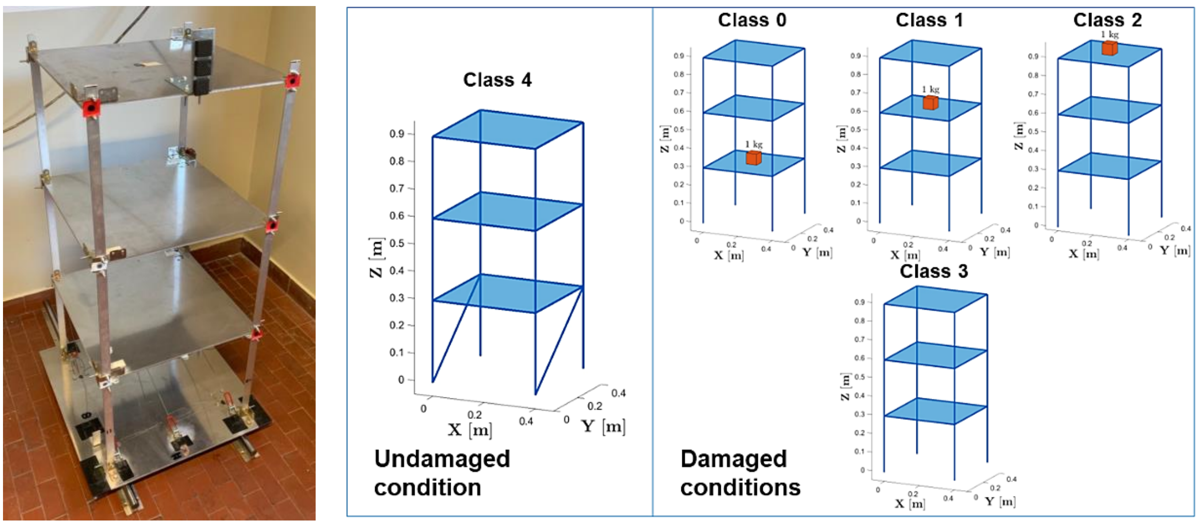
In this paper, the target system is an aluminum frame built in the Laboratory of Earthquake Engineering and Dynamics (EED Laboratory) at Politecnico di Torino. The structure is a three-story frame (Fig. 1(a)), which has a square plan of 0.4 m wide with a total height of 0.9 m. The structure is supported by four columns, fixed at the base end to a steel plate, which acts as a foundation plate. The majority of structural components, including columns and slabs, are made of aluminum. The slabs consist of square plate elements, each with a thickness of 5 mm. The floor decks are considered rigid within their plane and a diaphragmatic floor behavior is assumed for the frame. The columns, also including the diagonal braces, are constituted by rectangular sections 20x3 mm. Additionally, the braces are axial members pinned at the end nodes, which prevent the transfer of bending moments. Conversely, other connections, such as those between columns and slabs, can be assumed to be rigid. The mechanical characteristics of the aluminum structural elements are defined by a Young's modulus of 69000 MPa, a density of 2700 kg/m^3, and a Poisson's ratio of 0.326. A numerical model has been developed in MATLAB (Fig. 1(b)) with the aim of applying damages to the structure evaluating its modal behavior in terms of natural frequencies and modal shapes. In particular, the model has nine degrees of freedom, of which six in the horizontal plane (XY) and three rotations about the vertical axis (Z). The numerical structure is assumed to have 144 observations within the same structural condition, which are the experimental observations that occur in one day considering a 10-minute duration for each of these. A standard deviation of 690 MPa and 135 kg/m^3, respectively for Young's modulus and material density, are assumed to generate the noisy data, simulating the randomicity of the experimental observations.

2.2 Damage on diagonal bracings and mass variation over the three floors.

Several scenarios to simulate damage were applied both in the experimental and numerical frame, including a mass variation on floors (mass addition of 1 kg in the center of the slabs, considering it as a point load) and a damage to the braces located at the base. At first, modal parameters (natural frequencies and mode shapes) were experimentally derived by performing linear structural identification [12], with a total of 5 identifications for each damage class in addition to the undamaged one. The modes of vibration were then scaled with their respective natural frequency (Table 1). The same procedure was carried out on the numerical model which, however, offers the advantage of being able to derive many more data for each considered class. In this case, the ratio between the numerical and experimental features is 144:5. Here, a Gaussian Mixture (GM) model [13], defined as a multivariate (multidimensional), multi-component version of the classical Gaussian distribution, was used to compute distributions. Then, the Kullback-Leibler divergence (D_{KL}) (Eq. 6) [14] was considered to highlight the distance between the experimental distributions of the scaled eigenvectors and the numerical ones, to emphasize the correlation between the source and target data for the homogeneous TL application through TCA:

$$D_{KL}(p||q) = \sum_{x \in X} p_x \log_2 \frac{p_x}{q_x} \qquad 6$$

where p_x and q_x being two distributions, and $D_{KL}(p||q)$ representing the amount of information lost when q is used to describe p. Fig. 1(b) shows the undamaged (class 4) and damaged conditions: class 3 describes a fully-crack damage on both braces, while classes 0-1-2 also consider an added mass of 1 kg on the first, second and third floor, respectively. In Fig. 2(a) the scaled eigenvectors numerical distributions of class 4 and 3 are shown. The axes represent the first (x displacement of the first floor), fourth (y displacement of the first floor) and third (x displacement of the third floor) degrees of freedom. In Fig. 2(b), the D_{KL} of the experimental scaled eigenvectors of class 0, compared to the numerical model, is pointed out: each step corresponds to an addition of mass on the first floor equal to 0.1 kg on the numerical frame representing the structural condition of class 3, and it can be noted that the minimum divergence value is reached at step 10, equivalent to a mass addition of 1 kg (class 0), exactly corresponding to the added experimental mass. This procedure is performed also for classes 1-2-3 with the aim of corroborating the numerical model on the experimental identifications.

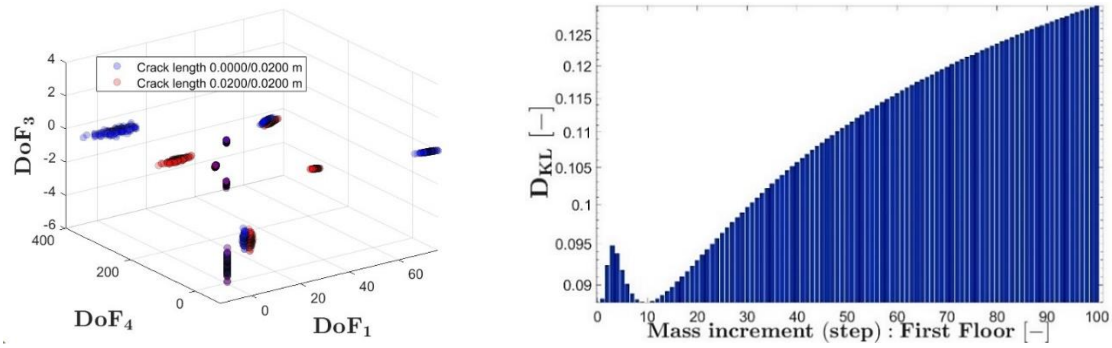


(a) (b)

Fig. 1. Three-story aluminum structure present at the Laboratory of Earthquake Engineering and Dynamics at Politecnico di Torino (a); Numerical representations of the three-story structure (classes 0-1-2-3-4) (b).

Table 1. First three experimental natural frequencies of the different structural conditions. Classes 0-1-2-3 describe the damage cases (mass variation and/or unbraced frame), while class 4 is the undamaged condition.

Class	0	1	2	3	4
f_{n1} [Hz]	3.10	3.05	2.93	2.81	4.18
f_{n2} [Hz]	6.31	6.34	6.33	5.93	6.39
f_{n3} [Hz]	8.08	8.09	8.23	8.13	11.87



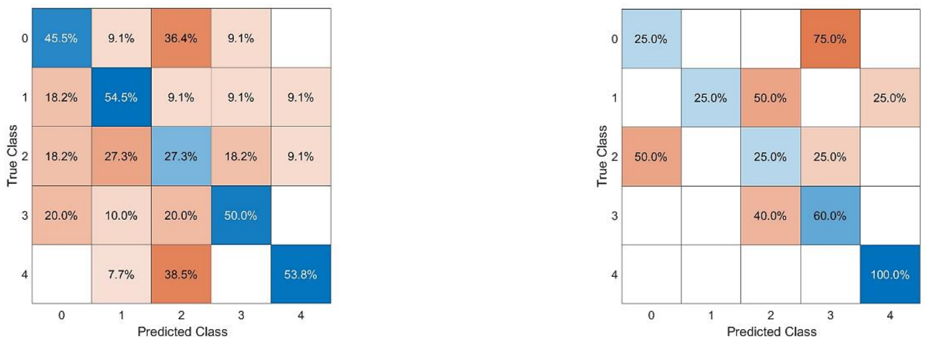
(a) (b)

Fig. 2. Numerical scaled eigenvectors distributions of class 4 (blue) and class 3 (red) (a); Kullback-Leibler divergence (D_{KL}) of the experimental scaled eigenvectors distribution (class 0) with respect to the numerical ones at each mass increment on the first floor (step = 0.1 kg) (b).

3. Results and discussion

3.1 Multi-class damage detection on the target structure

Once the damage classes were identified, it was possible to conduct a classification analysis. In this application, a fine k-Nearest Neighbor (k-NN) classifier 5-fold cross validated was utilized: this algorithm provides finely detailed distinctions between classes with a number of neighbors set to 1. Firstly, a model was trained and validated using the limited data from the target system (all classes), represented in this case by the available experimental data, consisting of the first 3 eigenvectors scaled with the corresponding natural frequencies. The target dataset is thus composed of 15 features having dimension 3 for each class (first 3 degrees of freedom), amounting to a total of 75 features (5 classes). Overall, the dataset was divided into 75% for training and validation, while the remaining 25% was used to test the reference model. The results in terms of model accuracy for multi-class damage detection in validation and testing are shown in Table 2. Additionally, the confusion matrices regarding the validation and testing (Fig. 3) results of the model are presented.



(a) (b)

Fig. 3. Validation Confusion Matrix (a) and Testing Confusion Matrix (b) for the model without TCA.

Table 2. Validation and testing accuracy of the model without TCA.

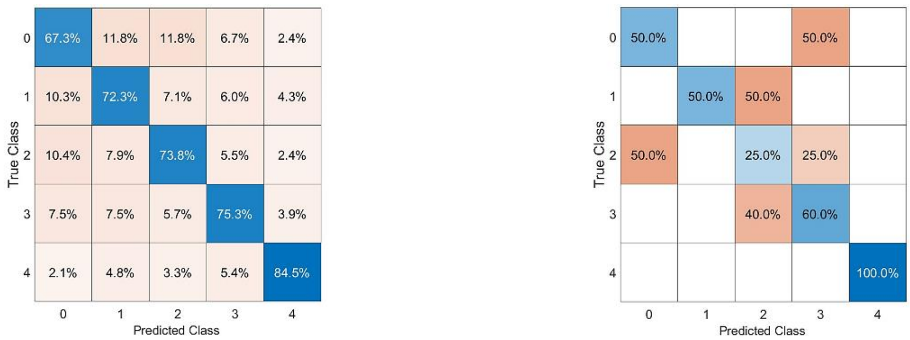
Model (k-NN)	Validation	Testing
Accuracy [%]	46.4	42.1

As we can observe, Fig. 3(a) reports the Validation Confusion Matrix for the model without TCA, while Fig. 3(b) shows the Testing Confusion Matrix for the model without TCA. In Fig. 3(a) and Fig. 3(b) the values correspond to the percentage rates of the classes prediction. On the left, labels describe the corresponding true classes (0-1-2-3-4), while on the bottom the predicted classes are shown; In general, the algorithm struggles to effectively classify the damage. In particular, the classes related to the addition of mass have a prediction accuracy of 25%, while there is an improvement for classes 3 and 4, reaching testing accuracies of 60% and 100% respectively (the latter corresponding to the undamaged state). This discrepancy can be attributed to the lack of data, which prevents target validation of the model, resulting in an accuracy of 46.4%. In the testing phase, this is reflected in the overall performance of the model, which achieves an accuracy of only 42.1%.

3.2 Transfer Component Analysis: multi-class damage detection on the target structure

Having observed that the classification algorithm does not perform optimally with only a few training data provided uniquely from the target system, the TL method was applied through TCA to merge the datasets related to the numerical model (source) (Fig. 1(b)) and the experimental one (target) (Fig. 1(a)). In general, several parameters need to be set during the domain adaptation phase using TCA. In this case, a linear kernel was used first, and a value of 0.1 was chosen to indicate the regularization/trade-off parameter μ, together with a transfer space dimension of 3. Classification was conducted using the same classifier (fine k-NN), but in this case, model validation relied on many more data, as the data transformed into the adapted domain, coming from both the source and the target system, were used. In total, a dataset of 75 (target) and 2160 (source) features are considered, maintaining an experimental to numerical ratio of 5:144. The model was trained and then validated with 75% of the available data, and then, to compare the results obtained without TCA, 25% of only target data were used to determine the accuracy of the model in testing.

Table 3 shows the testing accuracies of the target data of the model using TCA, while Fig. 4 reports again the Validation Confusion Matrix (Fig. 4(a)) and the Testing Confusion Matrix (Fig. 4(b)) for the model with TCA. The values correspond to the percentage rates of the class prediction for validation and testing. On the left, labels describing the corresponding true classes are shown, while on the bottom the predicted classes are described.



(a) (b)

Fig. 4 Validation Confusion Matrix (a) and Testing Confusion Matrix (b) for the model with TCA.

Table 3. Validation and testing accuracy of the model by applying TCA.

Model (TCA)	Validation	Testing
Accuracy [%]	74.6	52.6

The validation accuracy of the model has improved from 46.4% to 74.6% by integrating numerical data through TCA (Table 3): overall, it is noted that classes 3 and 4 exhibit the lowest predictive error percentage, while the effects caused by the addition of mass to the different floors (classes 0-1-2) are more challenging for the classification algorithm to be detected. Focusing on the testing data, it is possible to see how the accuracy has increased by applying TCA, rising from 42.1% to 52.6% (see Table 2 for reference). In particular, in Fig. 4, it can be observed that the confusion matrix partly mirrors that of the model validation, as the classes related to the further addition of mass return the lowest accuracy, while the undamaged state (class 4) is predicted without errors here as well.

4. Conclusions

In summary, the aim of this paper is to underline the potentiality of TL techniques in the SHM of multi-story buildings. In the work, a three-story aluminum frame has been chosen as target system, while its numerical model is considered as source system. The authors showed that the lack of experimental data in a multi-class damage detection strategy, causing little performance of the classification algorithm, can be solved exploiting numerical source datasets through TCA. The conclusions of the paper are reported hereinafter:

- The undamaged class is well predicted by both the strategies of classification, but the accuracy drops when the algorithm needs to predict the different damaged classes.
- The TCA was able to improve the accuracy of the multi-class damage detection task.

This represents a valuable support in the SHM framework.

References

[1] M. I. Friswell, "Damage identification using inverse methods," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1851), pp. 393-410.*, 2007.

[2] J. M. W. Brownjohn, J. Lee, and B. Cheong, "Dynamic performance of a curved cable-stayed bridge," *Engineering Structures, 21(11), 1015–1027*,1999,doi:https://doi.org/10.1016/S0141-0296(98)00046-7.

[3] Q. W. Zhang, T. Y. P. Chang, and C. C. Chang, "Finite-Element Model Updating for the Kap Shui Mun Cable-Stayed Bridge," *Journal of Bridge Engineering, 6(4), 285–293.*, 2001, doi: https://doi.org/10.1061/(ASCE)1084-0702(2001)6:4(285).

[4] K. Worden and G. Manson, "The application of machine learning to structural health monitoring," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1851), 515-537.* , 2006.

[5] K. Weiss, T. M. Khoshgoftaar, and D. D. Wang, "A survey of transfer learning," *J Big Data*, vol. 3, no. 1, Dec. 2016, doi: 10.1186/s40537-016-0043-6.

[6] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: a survey," *Journal of Machine Learning Research, 10(1), pp. 1633-1685.*, 2009.

[7] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans Neural Netw*, vol. 22, no. 2, pp. 199–210, Feb. 2011, doi: 10.1109/TNN.2010.2091281.

[8] S. J. Pan and Q. Yang, "A survey on transfer learning," *In IEEE Transactions on Knowledge and Data Engineering (Vol. 22, Issue 10, pp. 1345–1359).*, 2010, doi: https://doi.org/10.1109/TKDE.2009.191.

[9] C. T. Wickramarachchi, W. Leahy, K. Worden, and E. J. Cross, "On metrics assessing the information content of datasets for population-based structural health monitoring," in *European Workshop on Structural Health Monitoring: Special Collection of 2020 Papers-Volume 1*, 2021, pp. 494–504.

[10] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput. 10 (5), pp. 1299–1319.*, 1998.

[11] R. G. Congalton, Richard, Oderwald, and R. A. Mead, "Assessing Landsat classification accuracy using discrete multivariate analysis statistical techniques.," *Photogramm Eng Remote Sensing*, pp. 1671–1678, 1983, [Online]. Available: https://api.semanticscholar.org/CorpusID:56198043

[12] P. Van Overschee and B. De Moor, *Subspace identification for linear systems : theory, implementation, applications*. Dordrecht : Kluwer academic, 1996. [Online]. Available: http://lib.ugent.be/catalog/rug01:000736107

[13] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite Mixture Models," *Annu Rev Stat Appl*, vol. 6, no. 1, pp. 355–378, Mar. 2019, doi: 10.1146/annurev-statistics-031017-100325.

[14] S. Kullback and R. A. Leibler, "On Information and Sufficiency," *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951, doi: 10.1214/aoms/1177729694.