Challenges for SHM from structural repairs: an outlier-informed domain adaptation approach

P. Gardner, L.A. Bull, N. Dervilis, & K. Worden
Dynamics Research Group
Department of Mechanical Engineering, University of Sheffield,
Mappin Street, Sheffield S1 3JD, UK

Abstract

Data-based approaches to structural health monitoring are typically constructed on the assumption that the underlying data distributions in training will be the same as those experienced when the method is deployed. However, structural repairs alter the physical properties of the system, leading to a change in structural response. This change in response leads to a shift in the data distributions from the pre- to post-repair states — known as domain shift — invalidating the assumption that training, and subsequent operational data, come from the same underlying distribution. As a result, structural repairs represent a significant challenge to data-based approaches to structural health monitoring (SHM). Not only will domain shift cause an algorithm trained on the pre-repair data to fail to generalise, it will also make labels acquired from the pre-repair state redundant for building conventional data-based methods on the post-repair data. Transfer learning, in the form of domain adaptation, provides a solution to this problem, allowing knowledge from the pre-repair labels to be transferred to the post-repair dataset by forming a shared latent space where the pre- and post-repair dataset distributions are approximately equal. This paper presents a novel modification of a domain adaptation technique — joint domain adaptation — in creating *outlier-informed joint domain adaptation*, which can be used in transferring knowledge from pre- to post-repair states, forming a post-repair classifier that utilises all the pre-repair knowledge and generalises to post-repair data. The algorithm is demonstrated on an experimental dataset from a Gnat aircraft wing, where it is shown to outperform conventional data-based approaches and existing domain adaptation techniques.

Keywords: Transfer learning; domain adaptation; population-based structural health monitoring

1 Introduction

Performing structural repairs is an important aspect of asset management, and is crucial for keeping a structure in operation and ensuring safety. A consequence of repairing a structure is that the physical properties of a system are affected by these modifications, altering the system response. This change means that the pre-repair response of the system will be different to the post-repair response, causing a change in the underlying data distribution. This shift in distributions from pre-repair to post-repair is problematic for data-based approaches to structural health monitoring (SHM) [1], as conventional pattern recognition methods assume that the response of the structure under normal operation and in various health states will not change over time. This assumption, that the underlying data distributions from which samples are observed stays the same over time, means that a new data-based model must be formed once a structural repair has taken place, to ensure that the new system response behaviour is captured. As a consequence, all previously-labelled data points from the pre-repair state are now inapplicable to the post-repair model due to the data shift, meaning a new labelling campaign is required for the post-repair structure; this is often infeasible in industrial contexts. This paper proposes transfer learning, in the form of domain adaptation, as a solution to this problem, enabling pre-repair label data to be transferred to the post-repair dataset.

The aim of domain adaptation techniques is to manipulate the feature spaces such that a map is identified that projects the labelled source dataset (in this case the pre-repair dataset) and the unlabelled target dataset (here the post-repair dataset) onto a shared latent space where label information is transferable from source to target [2]. Domain adaptation has been explored in several SHM contexts: a population-based setting [3, 4], in fleet monitoring [5], transferring knowledge between operational conditions in condition monitoring [6, 7, 8] and in transferring

knowledge between sensors in a network [9]. However, none of these papers seek to overcome the problem of repair, which is the focus of this paper.

Several domain adaptation methods exist within the literature [10, 11, 12, 13, 14, 15, 16], which typically seek to reduce the distance between the source and target distributions in a shared latent space. One existing approach, *joint domain adaptation*, proposed by Long *et al.* [13], seeks to reduce the distance between the joint distributions in the shared latent space. This paper proposes a novel modification to the algorithm, improving the initial pseudo-label estimates in the target domain which help to define the target joint distribution. The modification introduced in this paper replaces a naïve self-labelling step with an outlier-informed procedure, forming *outlier-informed joint domain adaptation*. The approach is demonstrated on a Gnat trainer aircraft wing and is benchmarked against other domain adaptation techniques.

The outline of this paper is as follows. Section 2 introduces outlier-informed joint domain adaptation, outlining the novel modification to the initial pseudo-labelling step. The Gnat dataset is presented in Section 3, demonstrating the problem of repair. The dataset is subsequently used to demonstrate the effectiveness of domain adaptation, and in particular outlier-informed joint domain adaptation, for overcoming the problem of repair. Finally, a discussion and conclusions are presented in Section 4.

2   Outlier-informed joint domain adaptation

Domain adaptation is a subcategory of transfer learning that seeks to use data from *source* and *target* datasets in order to improve the predictive target function [2, 11, 17]. Typically, methods adapt the feature space, finding a nonlinear mapping onto a latent space where the source and target datasets lie on top of each other. Once in this latent space, a classifier can be trained on the source domain data that will generalise to the target domain data.

A challenge with any domain adaptation technique is that the target dataset is often unlabelled — hence the need for transfer learning. This poses challenges to knowledge transfer, as typically, domain adaptation techniques will try to match the joint distributions between the source and target datasets, i.e. $p(\boldsymbol{y}_s, X_s) \approx p(\boldsymbol{y}_t, X_t)$, and therefore target labelled observations are required. In order to overcome this problem, semi-supervised learning can be incorporated into the approach, where both labelled and unlabelled observations are used in training the predictive function [18]. In domain adaptation techniques, semi-supervised learning can be used to estimate *target pseudo-labels* $\hat{\boldsymbol{y}}_t$, which can be used to approximate the target joint distribution, i.e. $p(\hat{\boldsymbol{y}}_t, X_t) \approx p(\boldsymbol{y}_t, X_t)$ if $\hat{\boldsymbol{y}}_t \approx \boldsymbol{y}_t$.

Outlier-informed joint domain adaptation (O-JDA) seeks to obtain improved initial pseudo-label estimates by utilising outlier analysis techniques. The algorithm builds upon joint domain adaptation (JDA) proposed by Long *et al.* [13], replacing a naïve self-labelling step with pseudo-labels obtained from an outlier-assisted step.

2.1   Joint domain adaptation

Joint domain adaptation aims to find a nonlinear mapping ϕ onto a latent space where the distance between joint distributions for the source and target datasets are minimised [13]. This nonlinear transform moves the data into a reproducing kernel Hilbert space (RKHS), i.e. $\phi : \mathcal{X} \to \mathcal{H}$, where the aim is that $p(\boldsymbol{y}_s, \phi(X_s)) \approx p(\boldsymbol{y}_t, \phi(X_t))$. This goal requires knowledge of the target joint distribution, which is considered unlabelled. As a result, target pseudo-labels $\hat{\boldsymbol{y}}_t$ are obtained such that the target joint distribution can be approximated. In fact, due to the computational complexity in obtaining the full joint distribution, JDA instead finds an approximation, by minimising the distance between the marginal, $p(\phi(X_s))$ and $p(\phi(X_t))$, and class conditional distributions, $p(\phi(X_s) \mid \boldsymbol{y}_s = c)$ and $p(\phi(X_t) \mid \hat{\boldsymbol{y}}_t = c)$ where $c = \{1, \dots, C\}$ in the class label set $\mathcal{Y}$.

The distance criterion utilised in JDA is the (squared) maximum mean discrepancy (MMD) distance. The MMD distance assesses the difference between two distributions by measuring the distance between two empirical means, through a nonlinear mapping into an RKHS, specified by a kernel, $k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^{\mathsf{T}} \phi(\boldsymbol{x}_j)$ [19]. The MMD distance for the marginal distribution can be formed as,

$$D(p(X_s), p(X_t)) = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(\boldsymbol{x}_{s,i}) - \frac{1}{N_t} \sum_{i=1}^{N_t} \phi(\boldsymbol{x}_{t,i}) \right\|_{\mathcal{H}}^2 = \mathrm{tr}(KM) \tag{1}$$

where $K = \phi(X)^{\mathsf{T}} \phi(X) \in \mathbb{R}^{(N_s + N_t) \times (N_s + N_t)}$ given $X = X_s \cup X_t \in \mathbb{R}^{(N_s + N_t) \times d}$ where d is the dimension of the features. The empirical mean is defined though the matrix M as,

$$M(i,j) = \begin{cases} \frac{1}{N_s^2}, & x_i, x_j \in X_s \\ \frac{1}{N_t^2}, & x_i, x_j \in X_t \\ \frac{-1}{N_s N_t}, & \text{otherwise.} \end{cases} \tag{2}$$

In order to form an optimisation problem, JDA learns an empirical approximation of the optimal kernel by utilising a low-rank empirical kernel embedding $\tilde{K} = KWW^{\mathsf{T}}K$ [20]. Rewriting equation (1) using this low-rank embedding means the distance between the marginal distributions becomes,

$$D(p(X_s), p(X_t)) = \text{tr}(W^{\mathsf{T}}KMKW) \tag{3}$$

where $W \in \mathbb{R}^{(N_s+N_t)\times k}$ are now a set of weights which perform a reduction and transformation. It is noted that solving equation (3) as an optimisation problem is the same as performing transfer component analysis (TCA), which seeks to minimise the distance between marginal distributions only [10]. Equation (3) can be modified to include the distance between the class conditional distributions, approximating the distance between the joint distributions as,

$$D(p(\boldsymbol{y}_s, X_s), p(\boldsymbol{y}_t, X_t)) \approx \text{tr}(W^{\mathsf{T}}KM_cKW) \tag{4}$$

where $c = \{0, 1, \cdots, C\}$, which is the set of class labels in $\mathcal{Y}$, with the addition of zero such that when $c = 0$ the distance is equivalent to equation (3). The matrix M_c defines the empirical mean, but now considers the means of the marginal and class conditionals for each class c,

$$M_c(i,j) = \begin{cases} \frac{1}{N_s^{(c)} N_s^{(c)}}, & x_i, x_j \in \mathcal{D}_s^{(c)} \\ \frac{1}{N_t^{(c)} N_t^{(c)}}, & x_i, x_j \in \mathcal{D}_t^{(c)} \\ \frac{-1}{N_s^{(c)} N_t^{(c)}}, & \begin{cases} x_i \in \mathcal{D}_s^{(c)} x_j \in \mathcal{D}_t^{(c)} \\ x_j \in \mathcal{D}_s^{(c)} x_i \in \mathcal{D}_t^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

where $\mathcal{D}_s^{(c)} = \{\boldsymbol{x}_i : \boldsymbol{x}_i \in \mathcal{D}_s \wedge y(\boldsymbol{x}_i) = c\}$ are the instances that belong in class c given the true source label $y(\boldsymbol{x}_i)$ of $\boldsymbol{x}_i$ and $\mathcal{D}_t^{(c)} = \{\boldsymbol{x}_i : \boldsymbol{x}_i \in \mathcal{D}_t \wedge \hat{y}(\boldsymbol{x}_i) = c\}$ are the instances that belong in class c given the pseudo-target label $\hat{y}(\boldsymbol{x}_i)$ of $\boldsymbol{x}_i$ (where $\wedge$ is the logical AND symbol). The number of data points in each class for the source and target domains are therefore $N_s^{(c)} = |\mathcal{D}_s^{(c)}|$ and $N_t^{(c)} = |\mathcal{D}_t^{(c)}|$.

JDA forms the minimisation of the approximate distance between the joint distributions as the following regularised optimisation problem,

$$\min_{W^{\mathsf{T}}KHKW=\mathbb{I}} = \sum_{c=0}^{C} \text{tr}(W^{\mathsf{T}}KM_cKW) + \mu\text{tr}(W^{\mathsf{T}}W) \tag{6}$$

which is subject to the constraint, $W^{\mathsf{T}}KHKW = \mathbb{I}$ (kernel principal component analysis), removing the trivial solution $W = 0$; where $H = \mathbb{I} - 1/(N_s + N_t)\mathbf{1}$ is a centring matrix, $\mathbb{I}$ is an identify matrix and $\mathbf{1}$ a matrix of ones. The regularisation parameter μ controls a complexity penalty on the weights. The optimisation problem can be converted to an eigenvalue problem for W through a Lagrangian approach, where the eigenvectors required for the mapping correspond to the k smallest eigenvalues of,

$$\left(K \sum_{c=0}^{C} M_c K + \mu\mathbb{I}\right) W = KHKW\phi \tag{7}$$

where the k-dimensional transformed feature space is calculated by $Z = KW \in \mathbb{R}^{(N_s+N_t)\times k}$. It is noted that once observations are mapped into this new transfer space, a classifier trained on the source domain can be used to classify the target domain data given that the joint distributions are now 'close' given the MMD criterion.

A challenge for JDA is in acquiring the target pseudo-labels in order to obtain the target domain class conditional distributions. Long *et al.* propose a naïve self-labelling technique and iterating in an EM-like approach to refine the pseudo-label estimates [13]. This approach initialises W when M_c is calculated only for $c = 0$, meaning that only the marginal distributions are minimised initially, requiring no target labels. A classifier is constructed in the initial transfer space, trained on the source domain data and applied to the target data in order to obtain pseudo-label

estimates $\hat{\boldsymbol{y}}_t$. The algorithm uses these pseudo-labels in updating M_c for $c = \{0, 1, \ldots, C\}$, which are used in updating W and the cycle is repeated. However, if the initial pseudo-labels are poor, then the repeat iterations will compound the error, leading to a suboptimal mapping. O-JDA, therefore relaxes the assumption that marginal distribution matching provides good initial pseudo-label estimates, instead utilising an outlier analysis procedure defined in the next section.

2.2 Outlier-informed pseudo-labelling

O-JDA aims to improve the initial target pseudo-label estimates $\hat{\boldsymbol{y}}_t$, relaxing the assumption in JDA that the initial marginal distribution matching offers adequate initial guesses for the pseudo-labels [13]. This is performed through an outlier-informed step.

Outlier analysis seeks to detect whether an observation within a dataset is inconsistent with the rest of the data, using a discordancy measure, with the assumption that such data points are generated by some alternative mechanism compared to the rest of the data. The approaches have been well-studied within novelty detection for SHM [21, 22, 23, 1]. A commonly-used measure of discordancy for approximate Gaussian statistics is the Mahalanobis squared distance (MSD),

$$D_i^2 = (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}})^\mathsf{T} \hat{\Sigma}^{-1} (\boldsymbol{x}_i - \hat{\boldsymbol{\mu}}) \tag{8}$$

where $\boldsymbol{x}_i$ is the observation being considered, given a sample mean $\hat{\boldsymbol{\mu}}$ and covariance $\hat{\Sigma}$ determined from $X = \{\boldsymbol{x}_i\}_{i=1}^N$. This discordancy assumes the rest of the data are distributed given some normal distribution, i.e. $X \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ approximated by the sample mean and covariance. A threshold T for a given confidence bound can be set for an MSD discordancy [21], determining whether an observation is inlying or not.

Outlier analysis, however, suffers from the *curse of dimensionality*, meaning if the number of observations in X is small, i.e. $N < (d + 1)$ the sample covariance will be singular [24]. In this scenario outlier ensembles have been proposed [25], identifying an averaged discordancy from samples of the feature space. An outlier ensemble is formed from a committee of M models, each constructed from a sample mean $\hat{\boldsymbol{\mu}}_m$ and covariance $\hat{\Sigma}_m$, that provide a discordancy that can be used to form a weighted average discordancy measure $D_E^2(\boldsymbol{x}_i)$,

$$D_E^2(\boldsymbol{x}_i) = \frac{1}{M} \sum_{m=1}^M w_m D_m^2(\boldsymbol{x}_i) \tag{9}$$

where $D_m^2(\boldsymbol{x}_i)$ is the discordancy of the mth committee member with respect to the ith observation, constructed from $\hat{\boldsymbol{\mu}}_m$ and $\hat{\Sigma}_m$, with w_m being its associated weight — set in this paper to unity. Each member of the ensemble is formed from a random subset of features N_f from the d-dimensional set (where a guiding heuristic is that $N_f < \sqrt{N}$ [25]). This set of N_f features is used to calculate the sample mean $\hat{\boldsymbol{\mu}}_m$ and covariance $\hat{\Sigma}_m$ for the mth model, which in turn are used to calculate $D_m^2(\boldsymbol{x}_i)$ for all observations and averaged via equation (9) to find the ensemble discordancy.

The outlier-informed pseudo-labelling step, utilises a discordancy measure for each class, calculated from the labelled source dataset, as a method for estimating the target pseudo-labels. The discordancy for each class $D_c^2(\boldsymbol{x}_i^s) \, \forall \, c \in \{1 : C\}$ is calculated from a sample mean $\hat{\boldsymbol{\mu}}(\mathcal{D}_s^{(c)})$ and covariance $\hat{\Sigma}(\mathcal{D}_s^{(c)})$ obtained from the labelled source dataset $\mathcal{D}_s$. The main concept is that this discordancy, when applied to the unlabelled target observations $D_c^2(\boldsymbol{x}_i^t)$, will generate the lowest MSD values for target data from the same corresponding class, i.e. data from a class in the source domain will be most similar to data for its corresponding class in the target domain. In order to make objective comparisons between each class discordancy, given class imbalance, each discordancy is normalised by its threshold $\bar{D}_c^2(\boldsymbol{x}_i) = D_c^2(\boldsymbol{x}_i)/T$, where all discordancy values below T are set to zero. The normalised class discordancy measures
matrix of discordancy vectors for each class). From this discordancy feature space, target pseudo-labels are obtained by setting the label for the ith target instance to be the class c corresponding to the minimum discordancy, i.e. the c where $D_c^2 = \min(\bar{D}_C^2)$. This initial pseudo-labelling procedure is outlined in the first section of Algorithm 1.

The complete outlier-informed JDA algorithm, shown in Algorithm 1, has two main stages: obtaining initial pseudo-labels from the outlier approach, then performing JDA given these pseudo-label estimates. It is highlighted that the main difference to conventional JDA is that O-JDA allows the initial conditional distributions to be *more* different, given that the majority of the source and target data for each class are more similar to each other than any other class.

Algorithm 1 Outlier-informed joint domain adaptation

Initialise pseudo-labels

Set N_f and M ▷ Set the number of features and models
Form dataset $X = X_s \cup X_t$
Initialise $D_C^2 \leftarrow \{N \times C\}$
for $c = 1 : C$ **do**
    $\mathcal{D}_s^{(c)} = \{\boldsymbol{x}_i : \boldsymbol{x}_i \in \mathcal{D}_s \wedge y(\boldsymbol{x}_i) = c\} \, \forall i \in 1 : N$ ▷ Get source data for class c
    Initialise $D_M^2 \leftarrow \{N \times M\}$
    for $m = 1 : M$ **do**
        Sample N_f features from $\mathcal{D}_s^{(c)}$ ▷ Obtain random feature set
        Calculate sample mean $\hat{\boldsymbol{\mu}}_m$ and covariance $\hat{\Sigma}_m$ ▷ Obtain model parameters
        Calculate discordancy $D_m^2(X)$ ▷ Calculate model for all observations
        Store discordancy as column in D_M^2
    end for
    Calculate the ensemble average discordancy $D_c^2(X) = \text{average}(D_M^2(X))$
    Obtain the threshold T for a given confidence bound
    Normalise the discordancy $\bar{D}_c^2(X)$
    Store normalised ensemble average discordancy as column in $\bar{D}_C^2(X)$
end for
Obtain target pseudo-labels $\hat{y}(X_t) = c$ corresponding to $\min(\bar{D}_C^2(X_t))$

Domain adaptation

Set number of iterations
Construct $M_c \, \forall c \in 0 : C$ from equation 5 given $\{y(X_s), \hat{y}(X_t)\}$
for $i = 1 :$ iterations **do**
    Solve eigen-decomposition in equation 7 and select k smallest eigenvectors to construct W
    Calculate projection $Z = KW$
    Train classifier f on Z_s and $y(\boldsymbol{x}_i^s)$
    Update $\hat{y}(\boldsymbol{x}_i^t)$ from classifier f
    Update $M_c \, \forall c \in 1 : C$ from equation 5 given $\{y(X_s), \hat{y}(X_t)\}$
end for

---

3 Overcoming the repair problem on the Gnat aircraft wing

Structural repairs cause a shift in the underlying data distributions which is problematic for conventional machine learning approaches to SHM. Domain adaptation poses a potential solution, allowing label information to be transferred from the pre- to post-repair dataset. In the following section, a dataset from a Gnat trainer aircraft wing is introduced, demonstrating the challenge that structural repairs can cause, before demonstrating the effectiveness of domain adaptation, mainly in the form of O-JDA, as a potential solution.

The Gnat dataset was obtained from an experimental campaign that introduced pseudo-damage scenarios (as it was not possible to truly damage the structure) via the removal of nine inspection panels [22, 23, 26]. The response of the system, under a broadband white-noise excitation applied via an electrodynamic shaker below inspection panel four, was captured by a network of uni-axial accelerometers, forming transmissibility paths. A schematic of the aircraft wing is shown in Figure 1, highlighting the accelerometer, panel and transmissibility path locations. The transmissibility paths each covered a particular panel, i.e. the reference (denoted R) and target accelerometers form a path that crosses one particular inspection panel, where the panels and their associated transmissibilities are presented in Table 1. Each transmissibility covered a frequency range of 1024-2048Hz containing 1024 spectral lines, where the real and imaginary parts were converted into magnitude and phase, with the phase discarded. Damage was introduced sequentially, removing panels in each transducer group, A, B, and C, respectively. For each group, the panels were removed sequentially, i.e. P1, P2, then P3. In total 25 configurations were measured, each containing 100 repeat measurements, where each damage scenario was repeated twice for each transducer group, and normal condition data collected before and after. For more details about the experimental campaign, the interested reader is
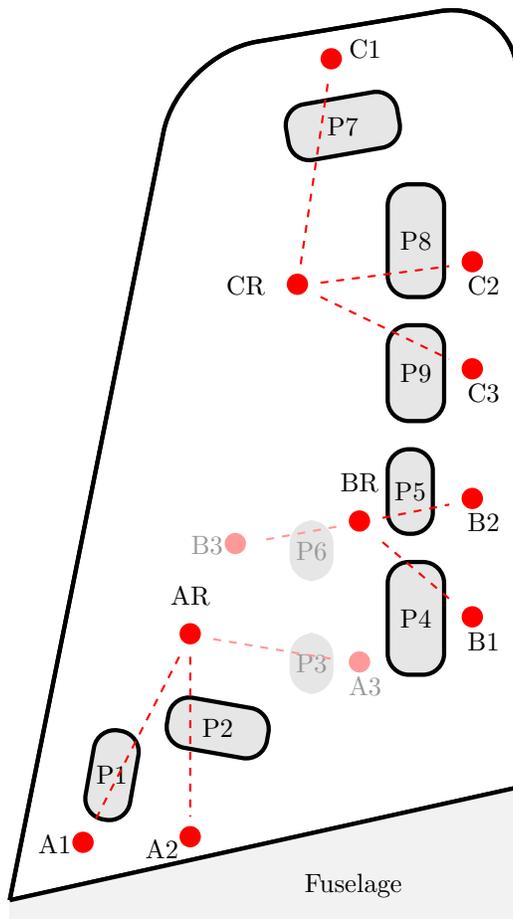
Figure 1: A representative schematic of the Gnat aircraft starboard wing (not to scale), indicating the locations of inspection panels, accelerometers and transmissibility paths. Recreated from [26].

Table 1: Accelerometers used to form transmissibilities and their associated inspection panel [26].

Panel	Associated transmissibility	Reference accelerometer	Response accelerometer
P1	T1	AR	A1
P2	T2	AR	A2
P3	T3	AR	A3
P4	T4	BR	B1
P5	T5	BR	B2
P6	T6	BR	B3
P7	T7	CR	C1
P8	T8	CR	C2
P9	T9	CR	C3

referred to [22, 23, 26].

The inspection panels are categorised into two types; a set of small panels {P3, P6} — both with an area of 0.00825m^2 [26] — and a set of larger panels {P1, P2, P4, P5, P7, P8, P9} — with areas greater than 0.0176m^2 [26]. In previous work, it was identified that the smaller panels caused confusion during localisation [26]. Furthermore, recent analysis on the dataset has demonstrated that a hierarchy of classifiers can be used to classify whether the class belongs to the large or small panel subsets, before localisation to each specific panel, with the approach showing boosted classification performance when compared to the complete problem [27]. For these reasons this paper considers the localisation problem considering only the large panels, i.e. $\mathcal{Y} \in \{1, 2, 4, 5, 7, 8, 9\}$. The feature space is formed
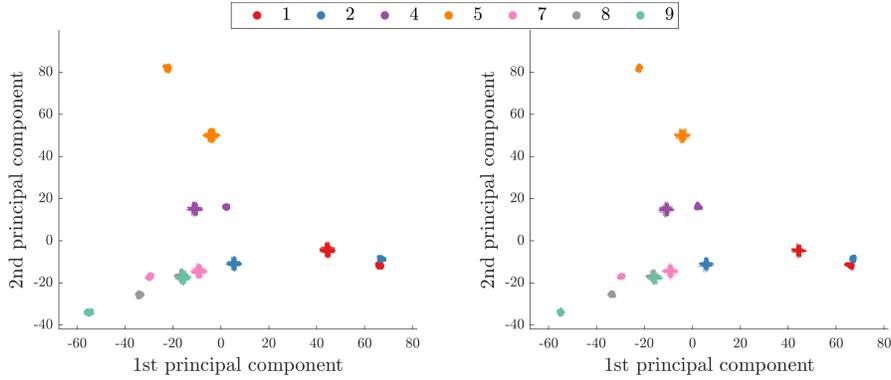
Figure 2: The first two principal components for the large panel feature set $\mathcal{X} = \log\{T_1, T_2, T_4, T_5, T_7, T_8, T_9\}$. Source data ($\bullet$) and target data ($+$) are depicted, where the left and right panels show the mapping training and testing data.

from the log transmissibilities that covered the large panels, i.e. $\mathcal{X} = \log(h + \{T1, T2, T4, T5, T7, T8, T9\})$, where a small constant h is added to shift the logarithm of zero elements [28].

The pre-repair (source) $X_s \in \mathbb{R}^{700 \times 7168}$, and post-repair (target) $X_t \in \mathbb{R}^{700 \times 7168}$, datasets were formed from the first and second repeat of each damage state. The panels were reattached with the same torque between the repeats. This scenario is reflective of repairs on operational structures, where inspection panels are removed and reattached as part of structural repairs. The source and target datasets are further divided into mapping training and testing sets, with 500 and 200 randomly-selected data points respectively. The first two principal components are shown in Figure 2, which demonstrates the data shift between the pre- (source) and post- (target) repair datasets. Clearly, a classifier trained on the pre-repair data will not generalise to the target domain.

Domain adaptation was applied to the large panel feature set in the form of O-JDA, JDA and TCA. Each algorithm was implemented with $\mu = 0.1$ and a flexible RBF kernel — where the hyperparameters were estimated using a median heuristic [19]. Two transfer components were extracted, highlighting a high level of dimensionality reduction. Each technique was trained using the (labelled) source and (unlabelled) target training datasets. O-JDA used an outlier ensemble approach with 1000 members ($M = 1000$) in each class discordancy and $N_f = 8$ (using the heuristic $N_f \approx \sqrt{N}$ [25]). A 99% confidence threshold was used to normalise each class discordancy measure, determined using a Monte Carlo approach with 10000 samples [21]. Figure 3 presents examples of the outlier analysis for classes one, two and five, which flag the corresponding class in the source testing dataset as inlying, and all other data as outlying due to the data shift, showing the need for domain adaptation. The spread of classes in the target data are generally similar to the class spread in the source data, and importantly the closest class for the target dataset in each ensemble corresponds to the class in which the ensemble was constructed, e.g. the ensemble for class one produces a discordancy measure that means class one is the closest class to the threshold for the target dataset. Classes two and four are the only classes where this behaviour is not distinct, causing overlap between two classes. Using the approach in Section 2.2, pseudo-labels are obtained from the unsupervised outlier-ensemble predictions. Two confusion matrices in Figure 4 illustrate the accuracy of these initial pseudo-labels, where a large amount of confusion has only occurred between classes one and two.

The feature spaces obtained from TCA, JDA and O-JDA are presented in Figures 5, 6 and 7. The TCA components have failed to bring the source and target dataset together due to the differences in the conditional distributions. This space is used to self-label the target domain in JDA, and hence causes JDA to find a mapping that produces negative transfer (see Figure 6), i.e. the wrong classes are matched. Due to the improved pseudo-label estimates, O-JDA produces a mapping with no negative transfer, correctly mapping the pre- and post-repair datasets onto a shared latent space. The performances of the KNN classifiers are shown in Figure 8 along with the classification performance of the outlier-informed pseudo-labels (OI) — the initial pseudo-labels used in O-JDA. O-JDA produces 100% classification accuracy, outperforming all other domain adaptation methods that have suffered from negative transfer.
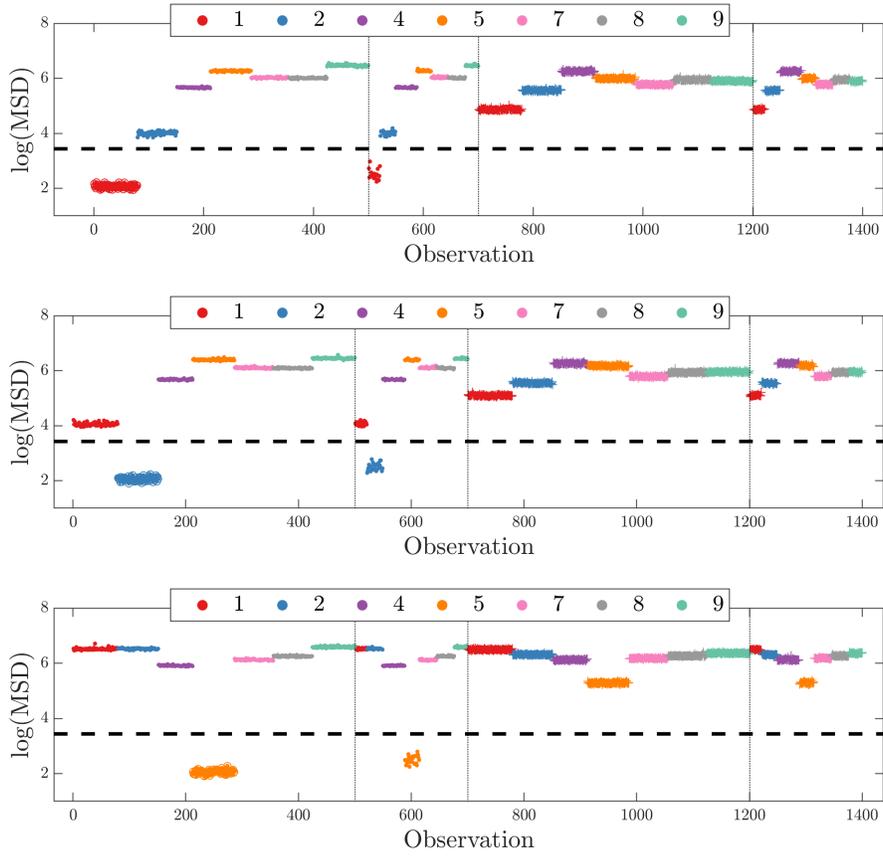
Figure 3: Outlier analysis results for the large panel dataset when considering class labels one (top panel), two (top middle panel) and five (bottom panel). Vertical lines denote the source training, source testing, target training and target testing datasets.
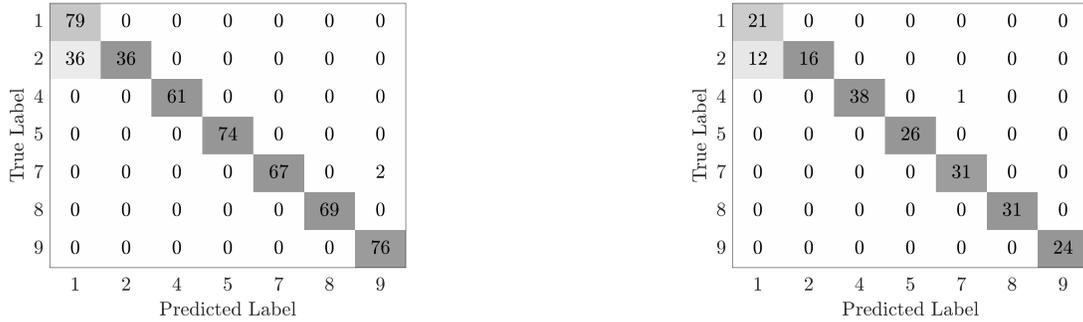

Figure 4: Confusion matrices for the target datasets when considering pseudo-labels from outlier analysis for the large panel dataset; mapping training, left panel and mapping testing, right panel.

4 Discussion and conclusions

Structural repairs cause shifts in the underlying data distributions, causing a classifier trained on the pre-repair data to fail to generalise to the post-repair data. Data shift is problematic for data-based approaches to SHM, and means that label data collected from the pre-repair structure cannot be used in classifying the post-repair structure. This paper has demonstrated that domain adaptation can be used as a solution to this problem, finding a shared latent space where the pre-repair data generalise to the post-repair data.

The paper introduces a novel modification of joint domain adaptation, named *outlier-informed joint domain*
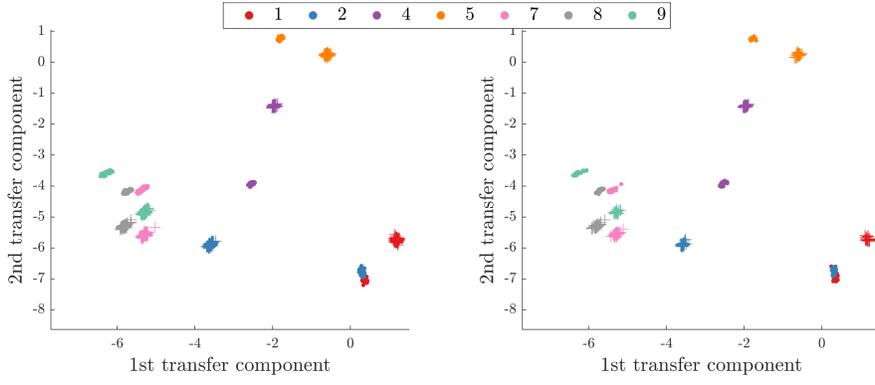
8

Figure 5: Transfer component analysis transfer components for the large panel feature set $\mathcal{X} = \log\{T_1, T_2, T_4, T_5, T_7, T_8, T_9\}$. Source data (•) and target data (+) are depicted where the left and right panels show the mapping training and testing data.
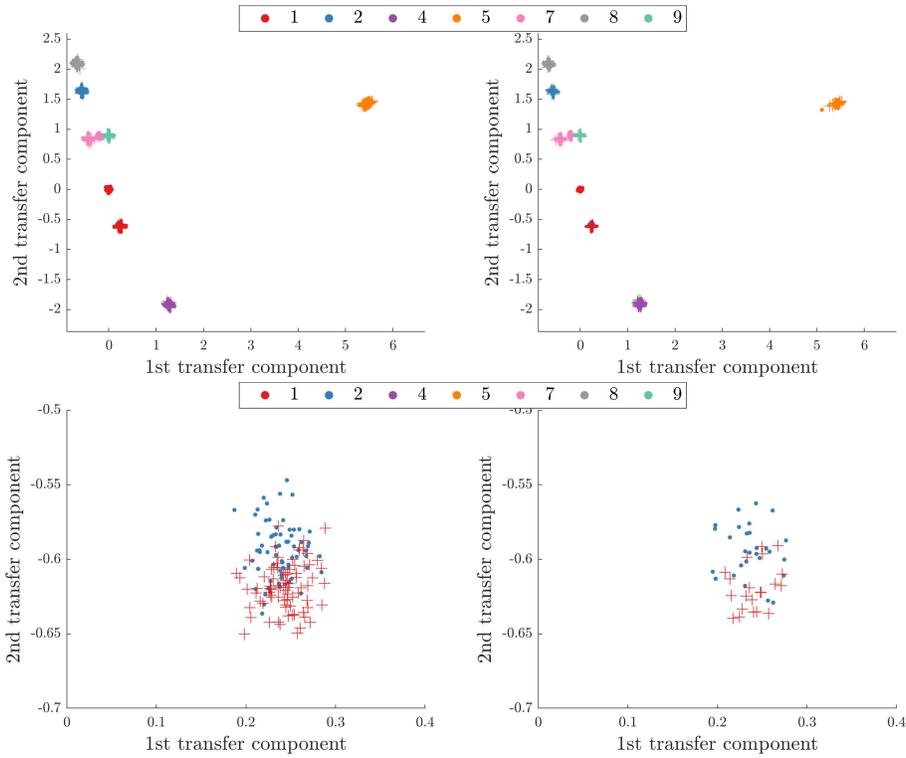

Figure 6: Joint domain adaptation transfer components for the large panel feature set $\mathcal{X} = \log\{T_1, T_2, T_4, T_5, T_7, T_8, T_9\}$. Source data (•) and target data (+) are depicted where the left and right panels show the mapping training and testing data. Top panels present the complete data space with the bottom panels showing a zoomed region where negative transfer has occurred.

adaptation. This algorithm uses an outlier analysis procedure for obtaining target pseudo-labels that can be used to approximate the post-repair (target) joint distribution. The approach showed a significant improvement on vanilla JDA, with O-JDA producing 100% classification accuracy on an experimental aircraft wing dataset, compared to JDA, which found a mapping that produced negative transfer. O-JDA has therefore been demonstrated to be an effective solution to overcoming the problem of repair.
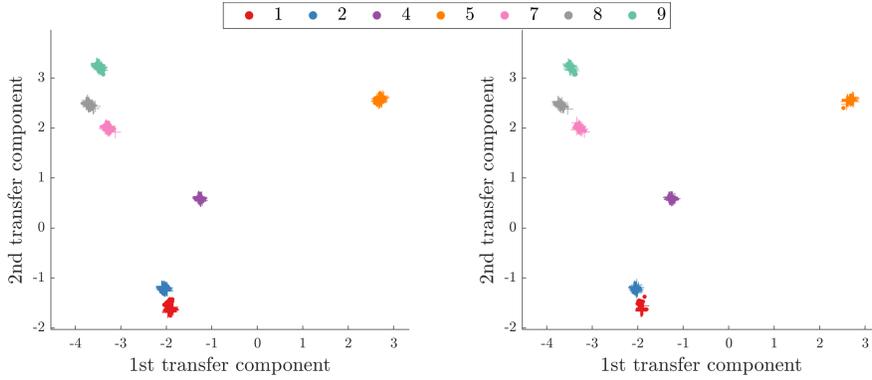
9

Figure 7: Outlier-informed joint domain adaptation transfer components for the large panel feature set $\mathcal{X} = \log\{T_1, T_2, T_4, T_5, T_7, T_8, T_9\}$. Source data ($\bullet$) and target data ($+$) are depicted where the left and right panels show the mapping training and testing data.
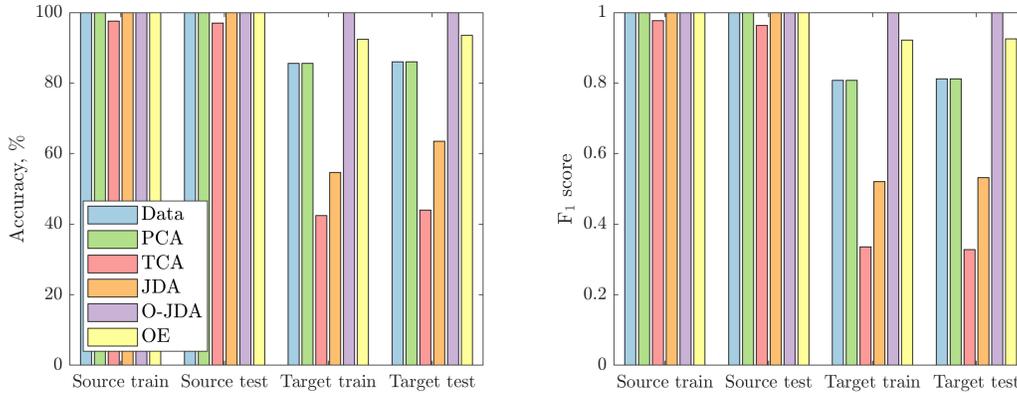

Figure 8: Comparison of classification performance given each feature space considering the large panel dataset; left panel, accuracy of predictions, right panel, F_1 scores. OI refers to the outlier-informed pseudo-labels.

Acknowledgements

References

[1] C. Farrar and K. Worden, *Structural Health Monitoring: a Machine Learning Perspective*. Chichester, UK: John Wiley & Sons, Ltd, 2012.

[2] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, pp. 1345–1359, 2010.

[3] P. Gardner, X. Liu, and K. Worden, "On the application of domain adaptation in structural health monitoring," *Mechanical Systems and Signal Processing*, vol. 138, p. 106550, 2020.

[4] P. Gardner, L. A. Bull, J. Gosliga, N. Dervilis, and K. Worden, "Foundations of population-based structural health monitoring, Part III: Heterogeneous populations, transfer and mapping," *Mechanical Systems and Signal Processing*, vol. 149, p. 107142, 2021.

[5] O. Fink, Q. Wang, M. Svensén, P. Dersin, W.-J. Lee, and M. Ducoffe, "Potential, challenges and future directions for deep learning in prognostics and health management applications," *Engineering Applications of Artificial Intelligence*, vol. 92, p. 103678, 2020.

[6] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, 2017.

[7] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-layer domain adaptation method for rolling bearing fault diagnosis," *Signal Processing*, vol. 157, pp. 180 – 197, 2019.

[8] Q. Wang, G. Michau, and O. Fink, "Domain adaptive transfer learning for fault diagnosis," in *2019 Prognostics and System Health Management Conference (PHM-Paris)*, 2019, pp. 279–285.

[9] D. Chakraborty, N. Kovvali, B. Chakraborty, A. Papandreou-Suppappola, and A. Chattopadhyay, "Structural damage detection with insufficient data using transfer learning techniques," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems*, 2011, p. 798147.

[10] S. Pan, I. Tsang, J. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, pp. 199–210, 2011.

[11] K. Weiss, T. Khoshgoftaar, and D. Wang, "A survey of transfer learning," *Journal of Big Data*, vol. 3, p. 29, 2017.

[12] L. Duan, D. Xu, and I. Tsang, "Learning with augmented features for heterogeneous domain adaptation," in *Proceedings of the 29th International Conference on Machine Learning, ICML 2012*, 2012.

[13] M. Long, J. Wang, G. Ding, J. Sun, and P. Yu, "Transfer feature learning with joint distribution adaptation," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 2200–2207.

[14] M. Long, J. Wang, G. Ding, S. Pan, and P. Yu, "Adaptation regularization: a general framework for transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, pp. 1076–1089, 2014.

[15] W. Li, L. Duan, D. Xu, and I. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 1134–1148, 2014.

[16] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, "Unsupervised domain adaptation for zero-shot learning," in *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2452–2460.

[17] O. Day and T. Khoshgoftaar, "A survey on heterogeneous transfer learning," *Journal of Big Data*, vol. 4, p. 29, 2017.

[18] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*. MIT press, 2006.

[19] A. Gretton, K. Borgwardt, M. Rasch, B. Schöolkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, pp. 723–773, 2012.

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

[21] K. Worden, G. Manson, and N. Fieller, "Damage detection using outlier analysis," *Journal of Sound and Vibration*, vol. 229, no. 3, pp. 647 – 667, 2000.

[22] K. Worden, G. Manson, and D. Allman, "Experimental validation of a structural health monitoring methodology: Part I. Novelty detection on a laboratory structure," *Journal of Sound and Vibration*, vol. 259, no. 2, pp. 323 – 343, 2003.

[23] G. Manson, K. Worden, and D. Allman, "Experimental validation of a structural health monitoring methodology: Part II. Novelty detection on a gnat aircraft," *Journal of Sound and Vibration*, vol. 259, no. 2, pp. 345 – 363, 2003.

[24] P. Rousseeuw and K. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, p. 212–223, 1999.

[25] L. Bull, K. Worden, R. Fuentes, G. Manson, E. Cross, and N. Dervilis, "Outlier ensembles: A robust method for damage detection and unsupervised feature extraction from high-dimensional data," *Journal of Sound and Vibration*, vol. 453, pp. 126 – 150, 2019.

[26] G. Manson, K. Worden, and D. Allman, "Experimental validation of a structural health monitoring methodology: Part III. Damage location on an aircraft wing," *Journal of Sound and Vibration*, vol. 259, no. 2, pp. 365 – 385, 2003.

[27] G. Tsialiamanis, D. Wagg, P. Gardner, N. Dervilis, and K. Worden, "On partitioning of an SHM problem and parallels with transfer learning," in *Proceedings of IMAC XXXVIII International Conference on Modal Analysis*, Houston, USA, 2020.

[28] S. R. Johnson and G. C. Rausser, "Effects of misspecifications of log-linear functions when sample values are zero or negative," *American Journal of Agricultural Economics*, vol. 53, no. 1, pp. 120–124, 1971.