Physics-Informed Transfer Learning in PBSHM: A Case Study on Experimental Helicopter Blades

J. POOLE, P. GARDNER, A. J. HUGHES, R. S. MILLS,
T. A. DARDENO, N. DERVILIS and K. WORDEN

ABSTRACT

Data for training Structural Health Monitoring (SHM) systems are often expensive or infeasible to obtain. Population-based SHM, which considers data across a population of structures, presents a potential solution to this issue. However, as differences between structures can lead to differing training and testing distributions, conventional machine learning methods may not generalise between structures. To address this issue, transfer learning (TL) can be used to leverage information across related domains. An important consideration when applying TL is how to asses similarity to identify and extract shared information. In unsupervised TL, a major challenge is that previous data-based metrics are limited to quantifying marginal distribution similarity in the unsupervised setting. This paper proposes utilising the Modal Assurance Criterion (MAC) between modes of healthy structures as a measure of data similarity to identify features that minimise conditional distribution shift. The MAC is incorporated into a feature selection criterion and a TL methodology is proposed. Moreover, the proposed methodology is shown to facilitate label sharing within a heterogeneous population of helicopter blades.

INTRODUCTION

In data-based SHM, diagnostics are often limited by the availability of data – particularly labelled data – as collecting labelled data for engineering applications is costly and often impractical. Population-based SHM (PBSHM) is a field that seeks to address this issue by considering data from a population of structures, thereby increasing the potential information available [1–3]. However, considering data from different structures invalidates the assumption that the training and testing data were drawn from the same distribution – an assumption made by conventional machine learning algorithms.

Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin St, Sheffield, S1 SJD UK

This issue motivates the application of transfer learning (TL), a field of machine learning that aims to address data scarcity issues by using labelled data from one or more source domains [4]. The assumption is that the source domain is related to the target domain, so knowledge transfer can improve performance in the target domain. Thus, TL algorithms aim to extract shared information by using limited target labels [5] and/or unlabelled target data [4, 6, 7]. Transfer learning has been successfully applied to the field of SHM for a number of applications [8–12].

A significant limitation of transfer learning approaches to SHM is the potential of performance degradation – a phenomenon called *negative transfer* [13]. Therefore, to avoid negative transfer, reliable metrics are needed to identify and extract shared information. However, sparse labels limit these metrics to measuring differences in the distribution of the unlabelled data (marginal-distribution shift) [14,15], which would not adequately quantify domain similarity if data generated by specific health-states followed different distributions (conditional -distribution shift).

To address these limitations, this paper proposes using the modal assurance criterion (MAC), between a source and target structure [16], only using data from the undamaged state. By incorporating the MAC into a feature-selection criterion, an unsupervised TL approach based on physics is shown to be able to extract shared information under conditional distribution shift. Furthermore, this approach only requires normal condition data in the target domain, so it can be applied before damage is observed in a target structure.

The structure of the paper is as follows. Section 2 introduces the core concepts of transfer learning and negative transfer, while Section 3 introduces the MAC-based feature selection criterion proposed in this paper. Section 4 introduces the experimental case study, as well as the TL methodology and shows the success of the approach by transferring label information between two dissimilar helicopter blades. The paper concludes with a discussion on physics-based similarity for transfer learning in PBSHM.

TRANSFER LEARNING

In *unsupervised transfer learning*, a source domain $\mathcal{D}_s = \{\mathbf{x}_{s,i}, y_{s,i}\}_{i=1}^{n_s}$, with n_s source instances $\mathbf{x}_{s,i}$ each with labels $y_{s,i}$, and a target domain $\mathcal{D}_t = \{\mathbf{x}_{t,j}\}_{j=1}^{n_t}$ with n_t unlabelled target instances $\mathbf{x}_{t,j}$, are used to learn a classifier that generalises to the target domain. It is assumed that there are differences in the marginal distributions $P(X_s) \neq P(X_t)$, and/or the conditional distributions $P(\mathbf{y}_s|X_s) \neq P(\mathbf{y}_t|X_t)$. Thus, TL aims to learn a predictive function using labelled data from the source that can provide accurate predictions on target domain data.

The Problem of Negative Transfer

A critical challenge for unsupervised TL in SHM is performance degradation caused by transfer, referred to as *negative transfer* [13,17,18]. The success of knowledge transfer largely relies on two main factors: the joint distributions of the domains $P(X_s, \mathbf{y_s})$ and $P(X_t, \mathbf{y_t})$ are sufficiently related and, the TL algorithms ability to utilise the shared information [13, 17]. Without labels, both of these issues are generally addressed by using unsupervised metrics that quantify the distance between the marginal distributions;

two popular choices are the *maximum mean discrepancy* (MMD) [19] and the *proxy-$\mathcal{A}$ distance* [20].

Data-based metrics have several drawbacks, particularly in engineering scenarios, where data are often sparse and expensive to obtain and label. In an unsupervised setting, these metrics are limited to estimating marginal distribution discrepancy, which will not provide a robust measure of similarity if there are large differences in the conditional distributions. Furthermore, the available data in the target domain may only represent a small subset of the underlying distribution; for example, at the start of a monitoring campaign of a target structure, only normal condition data would be available, whereas the source may have a range of health-states to transfer. In such cases, the source and target label space would be a subset of the source label space ($\mathcal{Y}_t \subset \mathcal{Y}_s$), meaning the metrics would not indicate whether the underlying distributions differ, but rather that the available subset of the distributions differs.

PHYSICS-BASED FEATURE SELECTION FOR TRANSFER

The issues with unsupervised metrics motivates using physical insight to indicate feature similarity. One possibility would be to utilise the mode shapes obtained from an undamaged structure during a limited time period, as recently discussed in [21]. The mode shapes describe the relative modal displacement of the structure, allowing identification of areas with high sensitivity to local structural changes [22]; hence, they can be used to indicate which features would have a similar sensitivity to damage in a specific location. In addition, using only a limited sample of mode shapes, similarity quantification can be achieved before damage is observed in the target structure, which is conducive to real-time damage diagnostics.

A popular measure of mode shape similarity is the modal assurance criterion (MAC) [16]. The MAC is a normalised scalar product between each pair of modal vectors $\boldsymbol{\phi}_s^{(i)}$ and $\boldsymbol{\phi}_t^{(j)}$ from two modal matrices Φ_s and Φ_t, which in this paper relate to the source and target domains respectively. The scalars are then arranged into a MAC matrix, assuming real-valued modal vectors; it is given by,

$$\text{MAC}(i,j) = \frac{|\boldsymbol{\phi}_s^{(i)T}\boldsymbol{\phi}_t^{(j)}|^2}{\boldsymbol{\phi}^{(i)T}_s\boldsymbol{\phi}_s^{(i)}\boldsymbol{\phi}_t^{(j)T}\boldsymbol{\phi}_t^{(j)}} \tag{1}$$

where $\text{MAC}(i,j) \in [0,1]$, with 0 indicating no correspondence and 1 is complete correspondence. The MAC between the source and target mode shapes is computed, and expectation between mode shapes corresponding to pairs of source and target feature pairs is computed; the metric can therefore be given by,

$$d_{\text{MAC}}(\Phi_s, \Phi_t) = \frac{1}{d}\sum_{i,j\in\mathcal{I}} \text{MAC}(i,j) \tag{2}$$

where $\mathcal{I} = (\mathbf{v}_s, \mathbf{v}_t) \mid \mathbf{v}_s, \mathbf{v}_t \in \mathbb{R}^d$ represents the pairs of feature indices, where $\mathbf{v}_s, \mathbf{v}_t$ are vectors of integers representing the source and target indices respectively and d is the feature dimension so the metric is normalised $d_{\text{MAC}} \in [0,1]$; this metric is called the MAC-discrepancy.

Standard feature selection methods are typically concerned with alleviating issues related to high feature dimension [23]. This paper proposes a *transfer feature criterion* (TFC) that incorporates MAC-discrepancy into a feature selection criterion to address the challenge of selecting features with high cross-domain similarity. The source loss is included in the criterion to balance the trade-off between informative and domain-invariant features, which is a pertinent issue for TL [4]; the criterion to maximise is given by,

$$\mathcal{L} = -\frac{1}{n_s}\sum_{n=1}^{n_s} L(f_s(\mathbf{x}_{s,i}), y_s) + \lambda d_{\mathrm{MAC}}(\Phi_s, \Phi_t) - \mu C \tag{3}$$

where $L(\cdot)$ represents the loss for a source model, λ and μ are trade-off parameters, and C represents a constraint to prevent the trivial solution of selecting the same feature multiple times; the constraint is given by,

$$C = \sum_{i=1}^{d}\sum_{i\neq j}[\mathbf{v}_{s,i} = \mathbf{v}_{s,j}] + [\mathbf{v}_{t,i} = \mathbf{v}_{t,j}] \tag{4}$$

where $[\cdot]$ represents the Iverson bracket, which takes 1 if the values are equal, otherwise, it is 0. To ensure the most similar source and target features are in correspondence, the target features are selected as,

$$\mathbf{v}_t = \underset{j}{\mathrm{argmax}}\,\text{MAC(i,j)} \tag{5}$$

A search strategy is required to find a set of source indices. The problem presented in this paper is low-dimensional so a greedy search is conducted. However, feature selection is a combinatorial optimisation problem and an exhaustive search is infeasible for high-dimensional feature spaces; thus, the TFC may be used with heuristic optimisation strategies when applied in these scenarios [24].

CASE STUDY: HETEROGENEOUS PAIR OF HELICOPTER BLADES

To explore the application of using modal similarity to inform transfer, a case study consisting of two heterogeneous full-scale helicopter blades is presented. Specifically, the blades are from a Robinson R44 and a Gazelle helicopter. Both blades are similar in size and internal structure, suggesting there is potential to share information. Importantly, there are several discrepancies, motivating the application of TL; these differences are summarised in Table I.

TABLE I. SUMMARY OF THE KEY DIFFERENCES BETWEEN THE ROBINSON R44 AND GAZELLE BLADES.

	Material	Mass (kg)	Length (m)	Width (m)	Leading edge thickness (mm)	Trailing edge thickness (mm)
Metal blade	steel	26.95	4.88	0.26	32.70	4.30
Composite blade	carbon fibre	37.00	4.83	0.30	28.10	1.00

Modal testing was conducted on the blades in a free-free configuration, utilising electrodynamic shakers attached in the flapwise direction, applying a continuous pink

Figure 1. The experimental setup to perform modal testing on a metal (right) and composite (left) blade simultaneously.

noise random excitation up to 800Hz, with a decay rate of 3DB/Octave and a sample rate of 1600Hz. Data was collected via ten uniaxial 100 mV/g accelerometers, placed on the underside of each blade along the length, at positions corresponding to the same non-dimensionalised length and width. To mitigate noise effects, ten frequency domain averages were obtained. Testing was conducted on both blades simultaneously, assuring data from both blades corresponded to the same environmental conditions; the experimental set-up is shown in Figure 1.

Data were collected for five health-states, including the normal condition and four pseudo-damage states, relating to adding small masses. The added masses were positioned at standardised lengths and widths of the blades and the size of the mass was scaled to maintain a consistent ratio between the added mass and blade mass. As such, the locations of damage should correspond to similar points of a given mode shape and the extent of “damage" can be considered equivalent for both blades. A summary of the datasets is given in Table II, where $L*$ and $W*$ refer to the non-dimensionalised length and width respectively, which were measured from the root and leading edge.

TABLE II. SUMMARY OF THE BLADE DATASETS. THE MASS RATIO BETWEEN THE METAL AND COMPOSITE BLADE IS 0.728.

Mass state	Repeats	Mass location ($L*$, $W*$)	Metal blade mass (g)	Composite blade mass (g)	Mass Ratio
M0	25	-	-	-	-
M1	10	(0.627, 0.577)	76.6	105.8	0.724
M2	10	(0.876, 0.577)	76.6	105.8	0.724
M3	10	(0.627, 0.577)	250.0	350.0	0.714
M4	10	(0.876, 0.577)	250.0	350.0	0.714

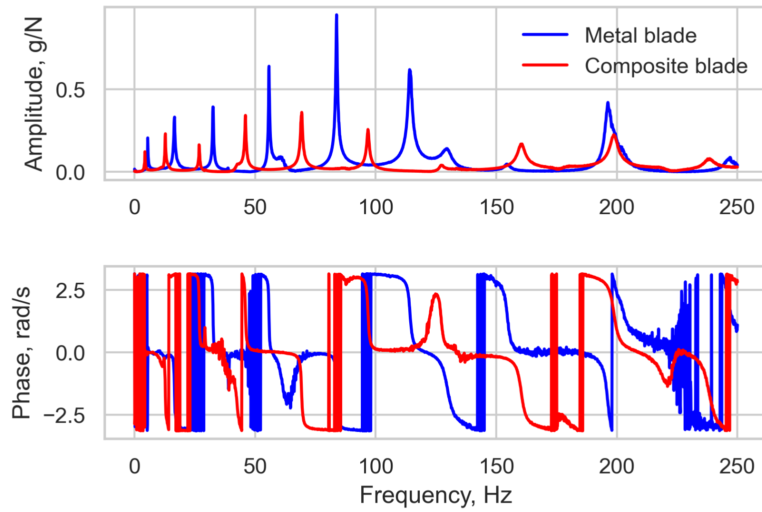


Figure 2. FRFs for the metal blade (blue) and composite blade (red) with no added masses.

The FRFs up to 250Hz from the sensor closest to the tip are presented in Figure 3. Initially, it can be seen that there are significant differences between the response of the blades; the peaks are shifted and there are differences in peak amplitude.

Transfer Learning

The objective was to classify the normal and four damaged states of the blades by transferring the acquired knowledge from one blade to another. This was accomplished by considering two tasks, wherein each blade was considered as both the source and the target domain. These tasks will be referred to as M→C when considering the metal blade as the source and composite blade the target, and C→M for the opposite case.

The FRF amplitude from one sensor close to the tip was used as a feature. Modal analysis was performed three times on the normal condition data to identify the regions of the FRF that should be put in correspondence for transfer. A subset of the frequencies was chosen by selecting a window of 20 features centred around the natural frequencies, assuming these regions predominantly correspond to the respective mode; an example of this feature space is shown in Figure 3. The MAC was found between the mode shapes of the blades to inform feature selection; the MAC matrix is given in Figure 4. Nine modes were identified in this range in the composite blade, but to benchmark the TFC against methods that require a homogeneous feature space, the 9th mode identified in the composite blade was removed. Note that the first eight modes were already in correspondence, although this may not always be the case, and in these scenarios using the MAC to bring modes into correspondence may be even more critical. Also, the last two modes have lower amplitude, potentially because excitation was applied close to a node in these modes.

A methodology utilising the TFC in conjunction with DA that further reduces domain discrepancy and feature dimension was implemented. First, a method that aligns the mean and standard deviations of the normal conditions, normal condition alignment (NCA) [25], was applied to account for discrepancies in amplitude. In addition, initial

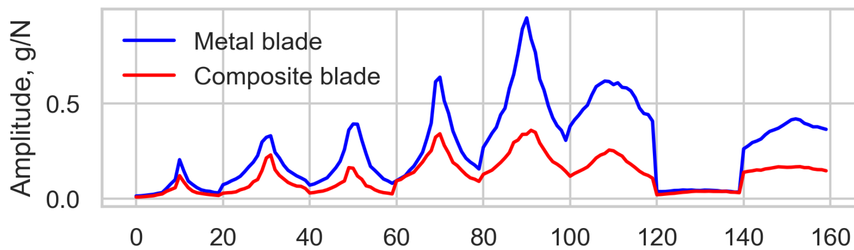


Figure 3. Example of the feature space after selecting a window of 20 frequencies centred around the natural frequencies for the metal blade (blue) and composite blade (red).

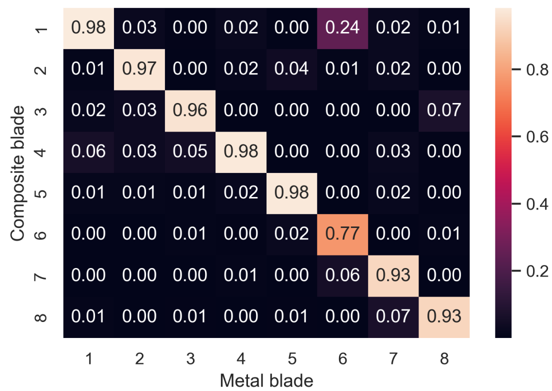


Figure 4. Modal assurance criterion matrix between the modes of the metal blade and composite blade normal condition.

alignment was found to be an important preprocessing procedure [25]; thus, NCA was applied prior to the application of additional DA algorithms.

Two kernel-DA algorithms that have been previously used in SHM were selected to benchmark the feature selection approach – transfer component analysis (TCA) [6] and balanced distribution adaptation (BDA) [26]. These algorithms project data into a shared latent space, by MMD-regularised kernel principal component analysis (PCA). TCA uses the conventional formulation of the MMD, whereas BDA also attempts to minimise the MMD between the class-conditional distributions $P(\mathbf{y}_s|X_s)$ and $P(\hat{\mathbf{y}}_t|X_t)$, where $\hat{y}_t$ are label predictions. These algorithms were also applied without the application of the TFC, in order to benchmark its performance. Finally, a K-nearest neighbours classifier (KNN) was applied for all methods, with one neighbour, as domains with low distribution divergence should be close in Euclidean space. The approach was also benchmarked against applying no additional TL beyond putting the features in correspondence - referred to as no TL.

Hyperparameter selection is challenging in unsupervised TL, as validation schemes typically rely on using labels. The regularisation hyperparameters were selected as $\lambda = 0.1$ for all cases, while the number of features for each method was determined using the source data, under the assumption that if the features are sufficiently discriminative in the source, they should also be in the target. This approach selects two features for all methods. Since the data were limited, accuracy was determined using leave-one-out

(LOO) validation; accuracy is given by,

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. An extension of the MMD that used labels, the joint-MMD (JMMD), was used to indicate distribution distance [27]. The JMMD was found on all the data since this metric requires a distribution of data. It should be noted that the limited sample size may impact the reliability of the JMMD.

Results

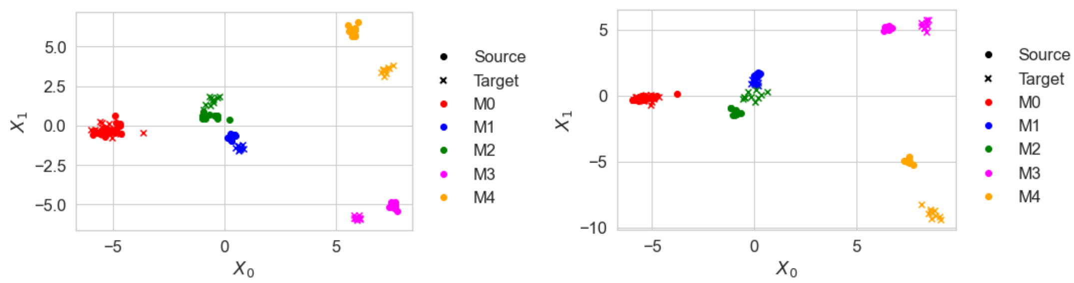
Table III shows the accuracy obtained from LOO validation and the JMMD. It can be observed that bringing features that relate to similar modes into correspondence allows some sharing of information, with NCA providing correct classification for some data. However, the accuracy is significantly compromised compared to supervised learning, shown by the source accuracy for each blade, which is indicative of domain shift.

Using the TFC, the frequencies relating to the fourth and fifth modes were selected, improving the generalisation of both NCA and the kernel DA methods. This result shows that comparing features that correspond to similar damage-sensitive modes can significantly reduce conditional distribution shift. It also validates the approach of scaling location and damage extent between structures with different geometries and material properties. The features are visualised using PCA in Figure 5. In this feature space, mass-states are in close correspondence and they are discriminative to the shift in normal conditions. However, since the minor damage classes are close, a small shift in the target led to a drop in classification performance.

Applying additional DA, specifically BDA (BDA+ in Table III), successfully reduced this domain shift in both tasks. This methodology resulted in a two-dimensional feature space, down from a high-dimensional raw FRF feature; the BDA features for both transfer tasks are presented in Figure 6. However, a potential limitation of BDA is that it assumes the label space is homogeneous, which may not always be the case in realistic scenarios; this issue requires further research into partial-DA algorithms [18].

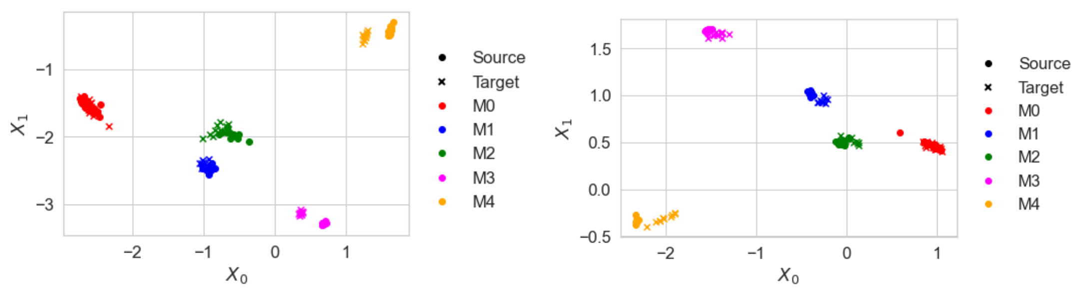
TABLE III. ACCURACY FOR THE SOURCE AND TARGET TEST DATA AND THE JMMD FOR THE TRANSFER METHODS APPLIED TO THE METAL AND COMPOSITE BLADE. TCA+ AND BDA+ ARE THE RESULTS OF APPLYING THESE ALGORITHMS AFTER THE TFC.

	No TL	NCA	PCA	TCA	BDA	TFC	TCA+	BDA+
M→C: Source Accuracy	0.98	0.98	0.98	1.00	1.00	1.00	0.98	1.00
M→C: Target Accuracy	0.15	0.71	0.72	0.49	0.77	0.85	0.88	1.00
M→C: JMMD	9.66	5.50	5.90	5.12	0.73	3.58	0.90	0.30
C→M: Source Accuracy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
C→M: Target Accuracy	0.38	0.58	0.58	0.85	0.85	0.89	0.85	1.00
C→M: JMMD	9.66	5.42	4.24	4.37	1.73	3.85	2.51	0.26



(a) (b)

Figure 5. PCA visualisation of the TFC-selected frequencies, corresponding to the fourth and fifth modes, for M→C (panel (a)) and C→M (panel (b)), representing 66% and 64% of the variance respectively.



(a) (b)

Figure 6. Features found via BDA applied to the TFC-selected frequencies, corresponding to the fourth and fifth modes, for M→C (panel (a)) and C→M (panel (b)).

DISCUSSION AND CONCLUSIONS

Negative transfer is a critical issue for TL in PBSHM. To reduce instances of performance degradation, TL algorithms should be able to reliably extract shared information while minimising the amount of discriminative information discarded. This paper suggests incorporating physical knowledge in the form of the MAC could present a solution to both of these issues in vibration-based SHM. Moreover, a methodology for TL that accounts for variations in geometry and material properties has been demonstrated to effectively transfer label information for damage classification across heterogeneous helicopter blades.

This paper proposes a method to evaluate how physics-based similarities could be used for transfer learning in SHM. However, to extend the findings, future work is required to investigate the requirements for sensor networks and to evaluate the approach in structures with realistic operating conditions. This includes robustness to noisy mode identification, influence of environmental effects, and nonlinearity.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the UK Engineering and Physical Sciences Research Council via grants EP/R006768/1, EP/R003645/1, EP/R004900/1 and EP/W005816/1. For the purpose of open access, the authors have applied for a Creative Commons Attribution (CC-BY-ND) licence to any Author Accepted Manuscript version arising.

REFERENCES

1. Bull, L.A., P. Gardner, J. Gosliga, T.J. Rogers, N. Dervilis, E.J. Cross, E. Papatheou, A.E. Maguire, C. Campos, and K. Worden. 2021. "Foundations of population-based SHM, Part I: Homogeneous populations and forms," *Mechanical Systems and Signal Processing*, 148:107141.
2. Gosliga, J., P. Gardner, L.A. Bull, N. Dervilis, and K. Worden. 2021. "Foundations of Population-based SHM, Part II: Heterogeneous populations – Graphs, networks, and communities," *Mechanical Systems and Signal Processing*, 148:107144.
3. Gardner, P., L. Bull, J. Gosliga, N. Dervilis, and K. Worden. 2021. "Foundations of population-based SHM, Part III: Heterogeneous populations – Mapping and transfer," *Mechanical Systems and Signal Processing*, 149:107142.
4. Zhuang, F., Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He. 2021. "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, 109(1):43–76.
5. Yosinski, J., J. Clune, Y. Bengio, and H. Lipson. 2014. "How transferable are features in deep neural networks?" *Advances in Neural Information Processing Systems*, 4(January):3320–3328, ISSN 10495258.
6. Jialin Pan, S., I.W. Tsang, J.T. Kwok, and Q. Yang. 2011. "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, 22(2).
7. Ganin, Y. and V. Lempitsky. 2015. "Unsupervised domain adaptation by backpropagation," *International Conference on Machine Learning*:1180–1189.
8. Xu, S. and H.Y. Noh. 2021. "PhyMDAN: Physics-informed knowledge transfer between buildings for seismic damage diagnosis through adversarial learning," *Mechanical Systems and Signal Processing*, 151:107374.
9. Gardner, P., L. Bull, J. Gosliga, J. Poole, N. Dervilis, and K. Worden. 2022. "A population-based SHM methodology for heterogeneous structures: Transferring damage localisation knowledge between different aircraft wings," *Mechanical Systems and Signal Processing*, 172:108918.
10. Bull, L., P. Gardner, N. Dervilis, E. Papatheou, M. Haywood-Alexander, R. Mills, and K. Worden. 2021. "On the transfer of damage detectors between structures: An experimental case study," *Journal of Sound and Vibration*, 501:116072.
11. Gardner, P., L. Bull, N. Dervilis, and K. Worden. 2021. "Overcoming the problem of repair in structural health monitoring: Metric-informed transfer learning," *Journal of Sound and Vibration*:116245.
12. Gardner, P., X. Liu, and K. Worden. 2020. "On the application of domain adaptation in structural health monitoring," *Mechanical Systems and Signal Processing*, 138:106550.
13. Zhang, W., L. Deng, and D. Wu. 2020. "Overcoming negative transfer: A survey," *arXiv*:1–15.
14. Ben-David, S., J. Blitzer, K. Crammer, and F. Pereira. 2007. "Analysis of representations for domain adaptation," *Advances in Neural Information Processing Systems*:137–144, ISSN 10495258, doi:10.7551/mitpress/7503.003.0022.

15. Gretton, A., B. Sriperumbudur, D. Sejdinovic, H. Strathmann, S. Balakrishnan, M. Pontil, and K. Fukumizu. 2012. “Optimal kernel choice for large-scale two-sample tests,” *Advances in Neural Information Processing Systems*, 2:1205–1213.
16. Allemang, R.J. 2003. “The modal assurance criterion–twenty years of use and abuse,” *Sound and vibration*, 37(8):14–23.
17. Wang, Z., Z. Dai, B. Póczos, and J. Carbonell. 2019. “Characterizing and avoiding negative transfer,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11293–11302.
18. Cao, Z., M. Long, J. Wang, and M. Jordan. 2018. “Partial transfer learning with selective adversarial networks,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*:2724–2732.
19. Gretton, A., K.M. Borgwardt, M. Rasch, B. Schölkopf, and A.J. Smola. 2007. “A kernel method for the two-sample-problem,” *Advances in Neural Information Processing Systems*:513–520, ISSN 10495258, doi:10.7551/mitpress/7503.003.0069.
20. Ben-David, S., J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J.W. Vaughan. 2010. “A theory of learning from different domains,” *Machine Learning*, 79(1-2):151–175, ISSN 15730565, doi:10.1007/s10994-009-5152-4.
21. Poole, J., P. Gardner, N. Dervilis, J.H. Mclean, T.J. Rogers, and K. Worden. 2023. “Towards Physics-Based Similarity Metrics for Robust Transfer Learning in PBSHM,” *To appear in the proceedings of IMAC-XL 2023*.
22. Farrar, C. and K. Worden. 2013. *Structural Health Monitoring: A Machine Learning Perspective*, Wiley, Chichester, ISBN 9781119994336.
23. Murphy, K.P. 2012. *Machine Learning: A Probabilistic Perspective*, MIT press.
24. Holland, J.H. 1992. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, MIT press.
25. Poole, J., P. Gardner, N. Dervilis, L. Bull, and K. Worden. “On statistic alignment for domain adaptation in structural health monitoring,” *Structural Health Monitoring*:14759217221110441.
26. Wang, J., Y. Chen, S. Hao, W. Feng, and Z. Shen. 2017. “Balanced distribution adaptation for transfer learning,” *Proceedings - IEEE International Conference on Data Mining, ICDM*:1129–1134.
27. Long, M., J. Wang, G. Ding, J. Sun, and P.S. Yu. 2013. “Transfer feature learning with joint distribution adaptation,” *Proceedings of the IEEE International Conference on Computer Vision*:2200–2207.