

# Domain Adaptation for Structural Health Monitoring

Ali I. Ozdagli and Xenofon Koutsoukos

*Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37212 USA*

*Ali.I.Ozdogli@vanderbilt.edu*

*Xenofon.Koutsoukos@vanderbilt.edu*

## ABSTRACT

In recent years, machine learning (ML) algorithms have gained significant interest in structural health monitoring (SHM) applications. Typical approaches assume the training and test data come from similar distributions. However, real-world applications, where an ML model is trained, for example, on numerical simulation data and tested on experimental data, are deemed to fail in detecting the damage, as the domain data are collected under different conditions and they do not share the same underlying features. This paper proposes to apply a domain adaptation approach for solving SHM problems where the classifier has access to the labeled training (source) and unlabeled test (target) domain data, and the source and target domains are statistically different. The proposed domain adaptation method seeks to form a feature space that is capable of representing both source and target domains by implementing a domain-adversarial neural network. This neural network uses H-divergence criteria to minimize the discrepancy between the source and target domain in a latent feature space. To evaluate the performance, we present two case studies where we design a neural network model for classifying the health condition of a variety of systems. The effectiveness of the domain adaptation is shown by computing the prediction accuracy of the unlabeled target data with and without domain adaptation. Furthermore, the performance gain of the domain adaptation over a well-known transfer knowledge approach called Transfer Component Analysis is also demonstrated. Overall, the results demonstrate that domain adaptation is a valid approach for SHM applications where access to labeled experimental data is limited.

## 1. INTRODUCTION

United States (US) has one of the most sophisticated infrastructures in the world (World Bank, 2019). However, according to a recent study conducted by the American Society of

Civil Engineers (ASCE), the US infrastructure is aging and failure on maintaining it may cost an economical loss in GDP as big as \$3.1 trillion (American Society of Civil Engineers, 2013, 2017). The condition of infrastructure for other modern societies is also under stress (Zachariadis, 2018). Overall, it is economically not viable to replace all deteriorating infrastructure due to limited resources, and the operations of maintenance, repair, and replacement should be prioritized accordingly. Acting proactively when a critical infrastructure requires care and preventing catastrophic damages call for novel and innovative approaches.

In the last few decades, structural health monitoring (SHM) has gained a lot of momentum as a means of detecting and localizing damages (Sohn, Farrar, Hemez, & Czarnecki, 2002). The introduction of machine learning (ML) into SHM enabled further refinement as mature pattern recognition techniques provide higher accuracy in recognizing structural damages compared to traditional methods (Farrar & Worden, 2012). Among many ML applications, supervised methods are of particularly useful (Kiranyaz et al., 2019). Especially, when coupled with artificial neural networks, supervised learning offers promising results for damage detection and localization (Park, Kim, Hong, Ho, & Yi, 2009; Dackermann, Li, & Samali, 2013; Nick, Asamene, Bullock, Esterline, & Sundaresan, 2015).

A majority of supervised SHM applications assume that the data used for training the damage condition classifier has the same distribution as the testing data. However, this assumption is problematic. First, it is unrealistic that one can obtain data belonging to a particular damage condition without actually harming the integrity of the structure before its service (Lu et al., 2016; Gardner, Liu, & Worden, 2020). In other words, creating labeled data based on the original state of the structure is not practical for supervised learning models. On the other hand, we can generate a labeled data set using a representative finite-element model or a similar scaled structure where introducing damages is a more cost-effective approach. The collection of labeled normal and damaged state data from this representative structure is called *source domain* and could be used for training a robust damage condition classifier. The

---

Ali I. Ozdagli et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

second problem with the supervised ML applications is that a model trained with labeled source domain data may fail to predict the condition original structure during testing time by looking at the unlabeled examples. The features for the original structure establish the *target domain*. Both source and target domains are distinct in a way that they have probability distributions which diverge from each other. To summarize, source domain is the model trained on labeled data derived from a representation of the original structure. The model trained on the unlabeled data directly sought from the original structure is the target domain. Both domains have different statistics, which is known as domain shift. The objective of domain adaptation is to design a new learning architecture that generalizes the prediction over both domains (Goodfellow, Bengio, & Courville, 2016). This generalization is achieved by finding a mapping that can extract domain-invariant features. Eventually, this mapping is expected to improve the prediction accuracy for the target domain compared to an architecture that does not implement domain adaptation. In brief, transfer of knowledge gained from source domain to target domain is conceptualized as domain adaptation (see Figure 1).

First attempt for domain adaptation started by addressing the distribution shift between labeled training and unlabeled test data. For example, Kernel Mean Matching (KMM) aims to minimize the covariate distribution between two datasets in a higher feature space called Reproducing Kernel Hilbert Space (RKHS) by reweighing the sample data. As a result, KMM is capable of producing a mapping that can match the test data distribution in RKHS (Gretton et al., 2009). While KMM outperforms ordinary classifiers and regressors, the improvement is limited to covariate shift such that the conditional distribution remains same ( $P_{train}(y|x) = P_{test}(y|x)$ ) but input distribution shifts ( $P_{train}(x) \neq P_{test}(x)$ ) across both domains (Bouvier, Very, Hudelot, & Chastagnol, 2019).

Many domain adaptation problems are susceptible to dataset shift where  $P(Y|X)$  is not conserved between source and target domains to its highest degree (M. Wang & Deng, 2018; Wilson & Cook, 2020). Thus, reweighing algorithms are not always effective in such cases. Modern domain adaptation techniques focus on finding a common latent space (also known as domain-invariant feature space) that represents both source and target domains. For example, as an improvement to KMM, maximum mean discrepancy (MMD) metric is introduced to measure the divergence between distributions and to compute a function in RKHS to maximize the difference in expectations between two probability distributions (Borgwardt et al., 2006). A well-known transfer learning method, transfer component analysis (TCA) uses this MMD metric to minimize the maximum expected distribution shift between source and target domain (Pan, Tsang, Kwok, & Yang, 2010). Additionally, Lu et al. utilized MMD as a loss function for the training of neural networks to im-

prove the prediction over target data using both source and target data during training. Similarly, Sun and Saenko employed CORAL, a metric similar to MMD for domain adaptation of classification problems.

The new generation domain approaches exploit adversarial training to find domain-invariant features (Wilson & Cook, 2020). These approaches adopts the zero-sum game where a label classifier (the network that predicts correct label of an input whether it is coming from source or target domain) is trained to deceive a domain classifier (the network that predicts whether the input is source or target domain data). For instance, Domain Adversarial Neural Network (DANN) uses gradient reversal layer during back-propagation to reverse the domain classifier weight derivatives to maximize the domain confusion (Ganin et al., 2016). Adversarial Discriminative Domain Adaptation (ADDA) uses a two-step approach where the network is first pre-trained on source data and then a domain classifier is trained to learn target domain features. As an alternative to DANN-type of domain adaptation, domain mapping approach uses GANs to translate a sample data in target domain to source domain (Benaim & Wolf, 2017; Zhu, Park, Isola, & Efros, 2017). However, these applications are limited to visual domain.

This paper introduces an effective domain adaptation approach to address the distribution shift between source and target domain for supervised machine-learning-based SHM applications. More specifically, we utilize a domain adversarial neural network (DANN) approach to predict the damage condition of a structure operating under a target domain using both labeled source and unlabeled target domain data during training time. The main purpose of the DANN architecture is learning features that represent both source and target domains. To achieve this goal, DANN implements a multi-task topology that combines a regular feed-forward neural network (NN) based damage classifier using source data with a domain discriminator NN which utilizes source and target domain data. The domain discrimination component enables the feed-forward NN to extract latent features underlying both domains by minimizing H-divergence between domains.

To demonstrate the suitability of the DANN for SHM applications, the paper investigates two case studies. The first case study focuses on a gearbox system with different damage conditions operating under low- and high-loads. A DANN model is trained with labeled low-load and unlabeled high-load data to predict the damage condition for the high-load operation of the gearbox. Additionally, for this case, DANN is compared to a well-known transfer knowledge method, Transfer Component Analysis (TCA) to show the performance gain from DANN. In the second case, the effectiveness of the domain adaptation from the numerical model to experimental data is studied for a small-scale three-story structure. The numerical model of the structure is used to simulate var-

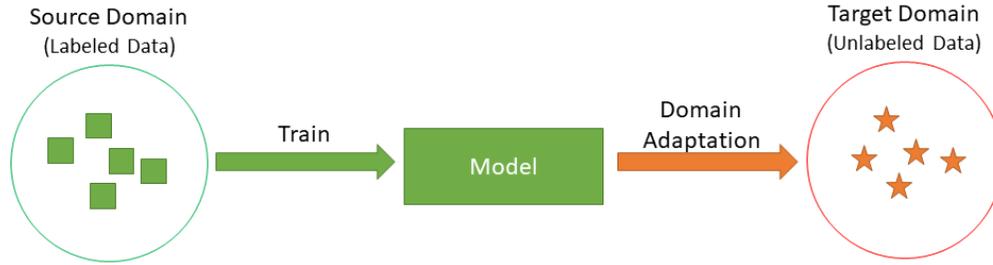


Figure 1. Concept of Domain Adaptation

ious damage conditions for the source domain whereas the experimental data constitutes the target domain. Results from both case studies indicate that domain adaptation is a viable method for SHM applications, and it increases the accuracy for damage condition prediction considerably. Additionally, the DANN can be considered as a potential ML architecture enabling appropriate knowledge transfer across the source and target domains.

For many machine-learning-based SHM applications focusing on damage detection and localization, a shift from source to the target domain is expected. Domain adaptation is a viable methodology for minimizing the distribution shift between source and target domains. This paper demonstrates that DANN is a suitable approach for learning latent features that underline both source and target domains. The case studies examined in this paper show that DANN improves the prediction accuracy of supervised damage detection and localization algorithms.

The rest of the paper is outlined as follows. First, Section 2 discusses condition monitoring briefly and formulates the domain shift problem. Section 3 introduces the DANN model for SHM applications. Section 4 presents case studies and the evaluation results. Lastly, Section 5 summarizes the paper and draw the conclusions.

The code to generate the results in this paper can be accessed from <https://github.com/aliirmak/DASHM>.

## 2. DOMAIN ADAPTATION IN SHM

In traditional SHM applications, vibration data is captured from various locations of the structure in the form of accelerations (Abdeljaber, Avci, Kiranyaz, Gabbouj, & Inman, 2017; Ozdagli & Koutsoukos, 2019). Meaningful features extracted from these measurements through time or frequency domain analysis establish the input space for a supervised learning model. Each data in the input space can be associated with a label describing the structural condition in terms of location of the damage and its intensity to form  $\{X, Y\}$ . Supervised learning algorithms require access to those labeled data for proper training. While the no-damage/normal data is often available when the structure is first erected, it is impractical

to abuse the structure just to obtain the data relevant to various damage conditions.

As a solution to the main fallback of the supervised learning methods, model-based SHM approaches exploit numerical models to establish a baseline for damage detection and damage localization (Mirzaee, Abbasnia, & Shayanfar, 2015; Figueiredo, Moldovan, Santos, Campos, & Costa, 2019). Numerical models can be useful for generating labeled source domain data. However, an ML model trained with source domain data may suffer from the uncertainty gap between the numerical model and the experimental structure (Catbas, Gokce, & Frangopol, 2013). Consequently, the learning model may not yield correct labels for the unlabeled target domain and may diagnose the damage improperly for the target structure. From the domain adaptation perspective, the distribution shift between source and target domain should be addressed (Singh, Azamfar, Ainapure, & Lee, 2020; Li, Li, He, & Qu, 2020). Accordingly, the problem for supervised SHM applications is finding domain-invariant features that represent both labeled source and unlabeled target domain.

In this paper, the source domain  $D_S$  consists of labeled data derived either from numerical simulations or from a particular state of the structure (for example, low wind, low traffic load, low-load, etc. corresponding to the normal operation). The target domain  $D_T$  is either the data captured from the experimental structure or an operational state of the structure that is not relevant to source domain (such as high wind, high traffic, high-load, etc. corresponding to stressing operations) and it is unlabeled. Then, the typical domain adaptation task for supervised SHM application is predicting the class for unlabeled target domain data using the knowledge gained from both source and target data.

For SHM, it is natural to consider a classification task where  $X = \{x_i\}_{i=1}^N$  is the input space of features and  $Y = \{y_i\}_{i=1}^N$  is the output space corresponding to the labels. Suppose that we have two different distributions over the  $\{X, Y\}$ : *i*)  $D_S$  is the source domain which contains the labeled source samples with  $S = \{(x_i, y_i)\}_{i=1}^n \sim D_S$ ; and *ii*)  $D_T$  is the target domain which consists of the unlabeled target samples with  $T = \{x_j\}_{j=1}^{n'} \sim D_T$ . We assume that the distributions for both domains are different such that  $D_S \neq D_T$ . This implies

that the distributions for the input space from  $S$  and  $T$  are not identical, namely  $p(X_S) \neq p(X_T)$ . Similarly, the conditional distributions that are used for inference may not match, that is  $p(Y_S|X_S) \neq p(Y_T|X_T)$ . Given  $D_S$  and  $D_T$ , the task for the domain adaptation is to build a classification model  $h(x)$  which can predict correct labels for samples from  $D_T$  using the knowledge learned from  $D_S$  and  $D_T$ .

### 2.1. Domain Adversarial Neural Network

A common domain adaptation approach is finding a mapping function that can minimize a probabilistic discrepancy metric between the two domains. The majority of these metrics focus on computing the divergence, i.e., the distance between two probability distributions. For example, the kernel mean matching (KMM) algorithm minimizes the mean distance in a kernel space by re-weighting the target domain with respect to source domain (Huang, Gretton, Borgwardt, Schölkopf, & Smola, 2007). The approach in (Sugiyama, Nakajima, Kashima, Buenau, & Kawanabe, 2008) proposes to minimize the Kullback-Leibler (KL) divergence for minimizing domain shifts. A well-known transfer learning algorithm called transfer component analysis (TCA) utilizes Maximum Mean Discrepancy (MMD) to minimize the distance between two domains in Hilbert space (Sejdinovic, Sriperumbudur, Gretton, & Fukumizu, 2013; Pan et al., 2010). Lastly, Ben-David et al. hypothesizes that a classifier-induced divergence, namely H-divergence is sufficient for domain adaptation.

H-divergence relies on distinguishing the examples of  $D_S$  and  $D_T$  and computing the domain divergence from the data in both domains. Accordingly, we label the data from  $D_S$  and  $D_T$  as 0 and 1, respectively. Then, we have a new dataset that can be described as:

$$U = \{x_i, 0\}_{i=1}^n \cup \{x_j, 1\}_{j=1}^{n'} \quad (1)$$

Then, the objective is to develop a function that predicts the class of the sample input  $\chi$  correctly, i.e.,  $f : \chi \rightarrow [0, 1]$ . Similarly,  $h'(x)$  is the learned model  $h' : \chi \rightarrow [0, 1]$ . Then, the generalized error is:

$$\epsilon = E[|h'(x) - f(x)|] \quad (2)$$

Given  $\epsilon$ , the H-divergence is approximately:

$$d = 2(1 - \epsilon) \quad (3)$$

One purpose of the domain adaptation is minimizing the H-divergence  $d$ . More details on the derivation of the divergence can be found in (Ben-David et al., 2010; Ganin et al., 2016).

There are inherently two tasks for implementing H-divergence based domain adaptation. First, we want to train a domain classifier  $h'(x)$  that can discriminate between source and target domains. At the same time, we want to

design a class predictor  $h(x)$  to correctly predict the class for the source domain data during training. It should be noted that the class predictor cannot be trained on target domain data as they are unlabeled. The ultimate aim of the domain adaptation is finding features that underline both the source and the target domain. Such a representation is expected to minimize the H-divergence and the domain predictor  $h'(x)$  should not be able to distinguish between the source and target domains.

The domain-adversarial neural network (DANN) approach introduced in (Ganin et al., 2016) exploits this objective by proposing a multi-task learning approach. The DANN is composed of three components: feature extractor, label predictor, and domain classifier (see Figure 2). The feature extractor (green colored) and label predictor (blue colored) layers are usually densely connected or convolutional layers. Both feature extractor and label predictor layers combined form a feed-forward neural network. This network uses only the labeled source data for training. The domain classifier (red-colored) is tasked with discriminating between the two domains. During the forward-propagation phase of the training, we can compute the loss over the labeled source data using the label predictor and the loss over both domain-labeled source and target data using the domain classifier. For typical applications, both losses can be logistic regression or cross-entropy functions depending on the ML task.

In back-propagation, a gradient reversal layer (denoted as GR in the figure) is added to the architecture to learn the latent features of both domains. This layer reverses the gradient after multiplying with a negative small constant. This negative gradient enforces the distribution of latent features extracted from both source and target domain to be indistinguishable. As a result, the entire network is expected to learn domain invariant features.

## 3. EVALUATION, RESULTS, AND ANALYSIS

For the evaluation of the proposed domain adaptation approach for SHM, two case studies are analyzed. The first case study investigates the prediction performance for the damage condition of a gearbox system under various torques. In the second case study, a three-story structure with several levels of damage conditions is used.

### 3.1. Case Study 1: Gearbox Fault Detection

#### 3.1.1. Dataset and Preprocessing

PHM Data Challenge 2009 introduced a dataset simulating various fault types for a generic gearbox system (*PHM data challenge 2009.*, 2009). Acceleration data were collected at the input and output shaft of the gearbox at different shaft speeds (30, 35, 40, 45, and 50 Hz) under two different loading conditions (low- and high-load). For each shaft speed

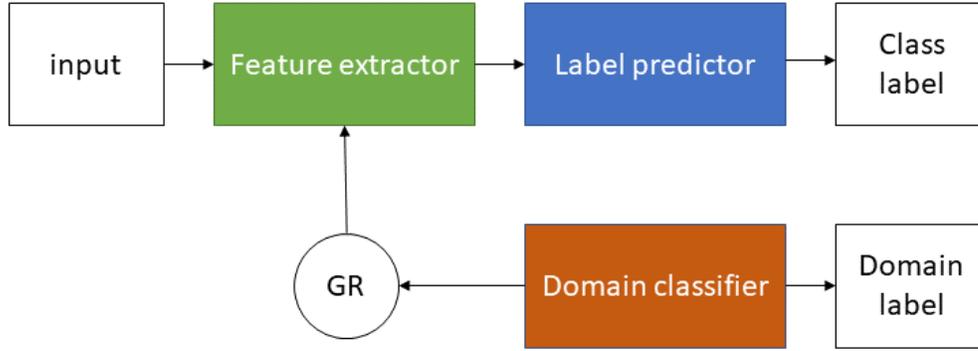


Figure 2. Simplified DANN architecture

and loading conditions, 6 fault types are simulated (normal, chipped gear tooth, broken gear tooth, bent shaft, imbalanced shaft, broken gear tooth with bent shaft). For each case which is the combination of fault type, shaft speed, and load condition, about 4 seconds of data is collected at a sampling rate of 66.67 kHz twice. *In this paper, only the output shaft vibration data is considered.*

According to the literature on gearbox fault detection (Chen, Li, & Sanchez, 2015; Jing, Zhao, Li, & Xu, 2017), the frequency domain provides a rich feature set for fault detection using vibration data. Thus, before training, all raw data is converted to the frequency domain using sliding-window Fast Fourier Transformation (FFT) also known as Short-Time Fourier Transform (STFT). The parameters for the transformations are selected as prescribed by the length of each window segment which is 1000 samples. The segments overlap by 80 percent and the sample length of FFT is 1200. The frequency resolution is  $\Delta f = 111$  Hz. After preprocessing, each damage condition case has about 2700 data points with 601 features per loading condition. The dataset is divided into source and target domains according to loading conditions. The source domain corresponds to low loading conditions consisting of all shaft speeds and fault types whereas the target domain is composed of the high-load operation. Since the task is detecting the type of the fault regardless of shaft speed, the data belonging to the same fault type are stacked together. Finally, both domain data is split into training and test data using a 4-to-1 ratio. All data is standardized with respect to the source training data and all labels are one-hot encoded.

### 3.1.2. Implementation

Three different models are developed: *Model 1*: source-only model which is trained only with source domain data; *Model 2*: the multi-tasking DANN model for training which uses both source and target domain data to discriminate the domain and predict the label; and *Model 3*: single-task DANN model for prediction and used only for testing. The architectures are shown in Figure 3. The source-only model is

a shallow network consisting of feature extraction (FE, colored in green) and class prediction (CP, colored in blue) layers. In addition to FE and CP layers, the multi-tasking DANN model includes the domain discriminator (DD, colored in red) layers and the gradient reversal (GR) layer. The single-task DANN model has the same structure with the source-only model but with updated weights where the FE contains the latent features that represent both source and target domains after training. Model 1 and Model 2 are trained using stochastic gradient descent. All the losses are chosen as categorical cross-entropy.

The low loading condition data represents the source domain whereas the high loading condition data corresponds to the target domain. During training, the DANN utilizes 128 data points (64 source and 64 target) per batch. We assume we have access to the source data labels but not to the target domain labels. The source (input, label) tuples are used explicitly for the class prediction task. For domain prediction, the source data is labeled as 0 and target data as 1, and then the labels are one-hot encoded. The domain predictor uses both domain data for training and creating domain invariant features. The source-only model is trained with 75 epochs whereas the DANN is trained for 200 epochs.

In addition to DANN, TCA is used for comparison. TCA utilizes training data from both source and target domain to realize dimension reduction using radial basis function as the kernel. After dimension reduction, an support vector machine (SVM) classifier is trained on the labeled source data. This classifier is also used to predict labels on unlabeled target data. Since TCA is essentially a set of matrix multiplications, the complete training dataset does not fit into the memory. Due to this limitation, only a quarter of training samples are used from both domains. TCA method is only applied to the first case and then discarded for the second case due to its low performance.

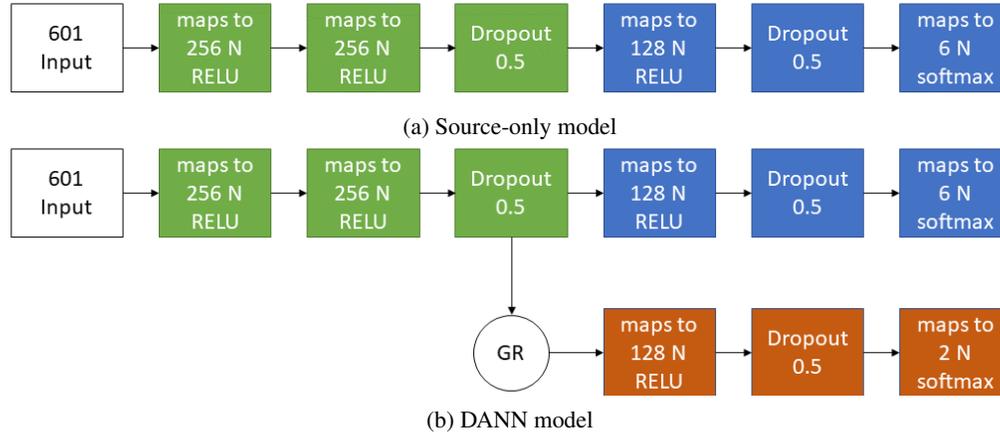


Figure 3. NN architectures for domain adaptation

### 3.1.3. Results

Table 1 shows the accuracy for source and target domain test data on the source-only model and the DANN model. The accuracy for predicting the source data is about 97 percent for both models. Without domain adaptation, the accuracy of the target data for the source-only model is 64 percent. The DANN improves the prediction on target data and increases the accuracy to 71 percent. The accuracy of the source-only and DANN models for the training data is 99.6 and 99.9 percent, respectively, such that the over-fitting is minimal. TCA method produces lower accuracy for source and target domain data ranging between 42 to 63 percent. DANN outperforms TCA significantly and this finding implies that TCA may not be used as a reliable domain adaptation method.

Table 1. Domain adaption performance for gearbox fault detection

Model	Input	Accuracy
Source-only Model	Source	97.51%
	Target	64.29%
DANN Model	Source	97.55%
	Target	71.79%
TCA Model	Source	61.87%
	Target	42.55%

## 3.2. Case Study 2: Structural Damage Detection

### 3.2.1. Structure and Numerical Model

This case studies the performance of domain adaptation when the training data are generated using a finite element model but the testing data are from an experimental structure. A small scale three-story structure is tested by Figueiredo, Park, Figueiras, Farrar, and Worden at the Los Alamos National Laboratory. The structure is excited with an electromagnetic shaker attached to its base. The accelerations at each floor

including the base are recorded at a sampling rate of 320 Hz for about 25 seconds. 7 damage conditions are considered where the stiffness of one or two out of four columns at different stories are reduced. Table-2) summarizes the damage conditions.

Table 2. Damage types for three story structure

Label	Damage Type
State #1	Baseline condition - Undamaged
State #2	87.5% stiff. red. in one column at first floor
State #3	87.5% stiff. red. in two columns at first floor
State #4	87.5% stiff. red. in one column at second floor
State #5	87.5% stiff. red. in two columns at second floor
State #6	87.5% stiff. red. in one column at third floor
State #7	87.5% stiff. red. in two columns at third floor

A simple lumped-mass numerical model of the structure is developed with the modal parameters identified by Figueiredo et al.. The error between numerical and experimental natural frequencies varies within the range of 0.4 to 3 percent. The numerical model is simulated at 320 Hz in MATLAB where the base is excited with the acceleration data captured at the base level from the experimental structure. Both numerical and experimental vibration data are converted to the frequency domain using STFT and the same parameters as in the first case study. For each damage condition, 2100 data points with 601 features are generated. After the data is split into training and testing, it is standardized and the labels are one-hot encoded.

### 3.2.2. Implementation

Similar to the first case study, three ML models are generated. For these three models, the data obtained using the numerical structural model constitutes the source domain and the data obtained using the experimental model represents the target domain. The topology and the parameters used in these models are the same as the one from the first Case 1. The source-

only domain is trained for 50 epochs and the DANN is trained for 200 epochs.

### 3.2.3. Results

Table 1 presents the performance of the DANN in predicting unlabeled target domain data. The prediction accuracy for the source data on the source-only and DANN model is almost 100 percent. When the domain adaptation is not considered, the accuracy of the target data is about 36 percent. With the DANN, the accuracy increases to 64 percent. The over-fitting for this case is small, as the accuracy of the source-only and DANN model for the training data is 100 and 99.98 percent, respectively.

Table 3. Accuracy for domain adaption from numerical to experimental world

Model	Input	Accuracy
Source-only Model	Source	99.92%
	Target	36.11%
DANN Model	Source	99.92%
	Target	63.08%

### 3.3. Discussion

To demonstrate the applicability of DANN, we consider two case studies. In the first case study, we predict the condition of a gearbox system running under high-load using the knowledge gained from low-load and high-load operation data. The second study focuses on transferring inference from labeled simulation data to unlabeled experimental data. While the improvement DANN provides for case 1 is modest, we observe a 30 percent increase in the target accuracy for case 2. It is clear that there is a big divergence between source and target domains for case 2. The learning model produced with the numerical data is not very successful in predicting correct labels for the target data without proper domain adaptation. However, the DANN is able to improve the accuracy of the target data by aligning features of source and target domain through H-divergence minimization. Specifically, for the first case, TCA produced low accuracy both for source and target domain data. This could be attributed to the fact that only a quarter of the total data set is used for the training since TCA is taxing to the memory similar to Principal Component Analysis for big number of samples. Thus, the generalization over both data set may be very well defined. Additionally, TCA uses SVM on dimension-reduced source domain dataset. SVM may not be the most suitable classifier for this application.

### 4. CONCLUSION

For many SHM methods based on supervised learning, experimental target data is often not available. For such cases,

a classification model trained with simulation data may not generate correct predictions for real data. Without addressing the data shift between the source and target domain, it is challenging to learn a model that can be used for SHM. This paper shows that domain adaptation is a viable approach to damage classification problems. Specifically, we show the applicability of adversarial domain adaptation using two case studies. In the first case, we study the fault detection performance for a gearbox system between low-load (source) and high-load (target) domains and we observed that the prediction accuracy improves using domain adaptation. Additionally, we compared DANN to TCA to demonstrate the performance gain from DANN over TCA. The second case focuses on detecting and locating damage for a three-story structure. Here, we utilized a numerical model of the structure for generating labeled source domain data and the experimental data for unlabeled target domain data. The results show that DANN increases classification performance significantly.

The current approach processes source and target data separately during training. In reality, for the majority of structural health monitoring applications, the structure is expected to be in healthy condition after the construction. As a result of this, target domain data labeled as *normal/undamaged* is accessible for training to some extent. For future research, novel domain adaptation methods should exploit this limited target domain data during training to extract more generalized latent features and to improve the adaptation. In addition, this paper uses only densely connected neural network architectures. There may exist a better representation mapping within different architectures utilizing convolution (Q. Wang, Michau, & Fink, 2019). Lastly, other domain adaptation strategies such as GAN-based discriminate approaches (Tzeng, Hoffman, Saenko, & Darrell, 2017) should be also explored.

### ACKNOWLEDGMENT

This work is supported in part by DARPA through contract number FA8750-18-C-0089. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the sponsor.

### REFERENCES

- Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, 388, 154–170.
- American Society of Civil Engineers. (2013). *Failure to act*. [https://www.asce.org/uploadedFiles/Issues\\_and\\_Advocacy/Our\\_Initiatives/Infrastructure/Content\\_Pieces/failure-to-act-economic-impact](https://www.asce.org/uploadedFiles/Issues_and_Advocacy/Our_Initiatives/Infrastructure/Content_Pieces/failure-to-act-economic-impact)

- summary-report.pdf. (Accessed: 2020-05-15)
- American Society of Civil Engineers. (2017). *2017 infrastructure report card*. <https://www.infrastructurereportcard.org/wp-content/uploads/2019/02/Full-2017-Report-Card-FINAL.pdf>. (Accessed: 2020-05-15)
- Benaim, S., & Wolf, L. (2017). One-sided unsupervised domain mapping. In *Advances in neural information processing systems* (pp. 752–762).
- Ben-David, S., Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., & Vaughan, J. W. (2010). A theory of learning from different domains. *Machine Learning*, 79(1-2), 151–175.
- Borgwardt, K. M., Gretton, A., Rasch, M. J., Kriegel, H.-P., Schölkopf, B., & Smola, A. J. (2006). Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14), e49–e57.
- Bouvier, V., Very, P., Hudelot, C., & Chastagnol, C. (2019). Hidden covariate shift: A minimal assumption for domain adaptation. *arXiv preprint arXiv:1907.12299*.
- Catbas, N., Gokce, H. B., & Frangopol, D. M. (2013). Predictive analysis by incorporating uncertainty through a family of models calibrated with structural health-monitoring data. *Journal of Engineering Mechanics*, 139(6), 712–723.
- Chen, Z., Li, C., & Sanchez, R.-V. (2015). Gearbox fault identification and classification with convolutional neural networks. *Shock and Vibration*, 2015.
- Dackermann, U., Li, J., & Samali, B. (2013). Identification of member connectivity and mass changes on a two-storey framed structure using frequency response functions and artificial neural networks. *Journal of Sound and Vibration*, 332(16), 3636–3653.
- Farrar, C. R., & Worden, K. (2012). *Structural health monitoring: a machine learning perspective*. John Wiley & Sons.
- Figueiredo, E., Moldovan, I., Santos, A., Campos, P., & Costa, J. C. (2019). Finite element-based machine-learning approach to detect damage in bridges under operational and environmental variations. *Journal of Bridge Engineering*, 24(7), 04019061.
- Figueiredo, E., Park, G., Figueiras, J., Farrar, C., & Worden, K. (2009). Structural health monitoring algorithm comparisons using standard data sets. *Los Alamos National Laboratory, Los Alamos, NM, Report No. LA-14393*.
- Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., ... Lempitsky, V. (2016). Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17(1), 2096–2030.
- Gardner, P., Liu, X., & Worden, K. (2020). On the application of domain adaptation in structural health monitoring. *Mechanical Systems and Signal Processing*, 138, 106550.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT press.
- Gretton, A., Smola, A., Huang, J., Schmittfull, M., Borgwardt, K., & Schölkopf, B. (2009). Covariate shift by kernel mean matching. *Dataset shift in machine learning*, 3(4), 5.
- Huang, J., Gretton, A., Borgwardt, K., Schölkopf, B., & Smola, A. J. (2007). Correcting sample selection bias by unlabeled data. In *Advances in neural information processing systems* (pp. 601–608).
- Jing, L., Zhao, M., Li, P., & Xu, X. (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement*, 111, 1–10.
- Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2019). 1d convolutional neural networks and applications: A survey. *arXiv preprint arXiv:1905.03554*.
- Li, J., Li, X., He, D., & Qu, Y. (2020). A domain adaptation model for early gear pitting fault diagnosis based on deep transfer learning network. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 234(1), 168–182.
- Lu, W., Liang, B., Cheng, Y., Meng, D., Yang, J., & Zhang, T. (2016). Deep model based domain adaptation for fault diagnosis. *IEEE Transactions on Industrial Electronics*, 64(3), 2296–2305.
- Mirzaee, A., Abbasnia, R., & Shayanfar, M. (2015). A comparative study on sensitivity-based damage detection methods in bridges. *Shock and Vibration*, 2015.
- Nick, W., Asamene, K., Bullock, G., Esterline, A., & Sundaresan, M. (2015). A study of machine learning techniques for detecting and classifying structural damage. *International Journal of Machine Learning and Computing*, 5(4), 313.
- Ozdagli, A. I., & Koutsoukos, X. (2019). Machine learning based novelty detection using modal analysis. *Computer-Aided Civil and Infrastructure Engineering*, 34(12), 1119–1140.
- Pan, S. J., Tsang, I. W., Kwok, J. T., & Yang, Q. (2010). Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2), 199–210.
- Park, J.-H., Kim, J.-T., Hong, D.-S., Ho, D.-D., & Yi, J.-H. (2009). Sequential damage detection approaches for beams using time-modal features and artificial neural networks. *Journal of Sound and Vibration*, 323(1-2), 451–474.
- Phm data challenge 2009*. (2009). <http://www.phmsociety.org/competition/09>. (Accessed: 2009-09-28)
- Sejdinovic, D., Sriperumbudur, B., Gretton, A., & Fukumizu, K. (2013). Equivalence of distance-based and rkhs-based statistics in hypothesis testing. *The Annals of*

- Statistics*, 2263–2291.
- Singh, J., Azamfar, M., Ainapure, A., & Lee, J. (2020). Deep learning-based cross-domain adaptation for gear-box fault diagnosis under variable speed conditions. *Measurement Science and Technology*, 31(5), 055601.
- Sohn, H., Farrar, C. R., Hemez, F. M., & Czarnecki, J. J. (2002). *A review of structural health review of structural health monitoring literature 1996-2001* (Tech. Rep.). Los Alamos, New Mexico: Los Alamos National Laboratory.
- Sugiyama, M., Nakajima, S., Kashima, H., Buenau, P. V., & Kawanabe, M. (2008). Direct importance estimation with model selection and its application to covariate shift adaptation. In *Advances in neural information processing systems* (pp. 1433–1440).
- Sun, B., & Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision* (pp. 443–450).
- Tzeng, E., Hoffman, J., Saenko, K., & Darrell, T. (2017). Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7167–7176).
- Wang, M., & Deng, W. (2018). Deep visual domain adaptation: A survey. *Neurocomputing*, 312, 135–153.
- Wang, Q., Michau, G., & Fink, O. (2019). Domain adaptive transfer learning for fault diagnosis. In *2019 prognostics and system health management conference (phm-paris)* (pp. 279–285).
- Wilson, G., & Cook, D. J. (2020). A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5), 1–46.
- World Bank. (2019). *2017 infrastructure report card*. <https://lpi.worldbank.org/international/global>. (Accessed: 2020-05-15)
- Zachariadis, I. A. (2018). *Investment in infrastructure in the eu: Gaps, challenges, and opportunities*. [https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS\\_BRI\(2018\)628245](https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2018)628245). (Accessed: 2020-05-15)
- Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223–2232).
- research scientist with the Institute for Software Integrated Systems, Vanderbilt University. His research work is in the area of cyber-physical systems with an emphasis on security and resilience, control, diagnosis and fault tolerance, formal methods, and adaptive resource management. He received the Ph.D. degree in electrical engineering from the University of Notre Dame. He is a Fellow of the IEEE.

## BIOGRAPHIES

**Ali I. Ozdagli** is a graduate student in the Department of Electrical Engineering and Computer Science at Vanderbilt University. Prior to this, he got his M.S. in civil engineering from the University of Notre Dame and his Ph.D. from Purdue University. His research focuses on adapting machine learning approaches to structural health monitoring applications.

**Xenofon Koutsoukos** is a professor with the Department of Electrical Engineering and Computer Science and a senior