Structural Damage Detection with Insufficient Data using Transfer Learning Techniques

Debejyo Chakraborty[†*], Narayan Kovvali[‡], Bhavana Chakraborty[‡], Antonia Papandreou-Suppappola[‡], and Aditi Chattopadhyay[§]

[†]Global Research & Development, General Motors Corporation, Warren, MI
[‡]School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ
[§]School of Mechanical, Aerospace, Chemical and Materials Engineering, Arizona State University, Tempe, AZ

ABSTRACT

The effective detection and classification of damage in complex structures is an important task in the realization of structural health monitoring (SHM) systems. Conventional information processing techniques utilize statistical modeling machinery that requires large amounts of 'training' data which is usually difficult to obtain, leading to compromised system performance under these data-scarce conditions. However, in many SHM scenarios a modest amount of data may be available from a few different but related experiments. In this paper, a new structural damage classification method is proposed that makes use of statistics from related task(s) to improve the classification performance on a data set with limited training examples. The approach is based on the framework of transfer learning (TL) which provides a mechanism for information transfer between related learning tasks. The utility of the proposed method is demonstrated for the classification of fatigue damage in an aluminum lug joint.

Keywords: structural health monitoring, damage classification, time-frequency analysis, matching pursuit decomposition, hidden Markov models, transfer learning

1. INTRODUCTION

A key task in the realization of structural health monitoring (SHM)[1–3] systems is the effective detection and classification of damage in complex structural components. A large number of information processing techniques have been investigated in the last few years, including joint time-frequency methods such as wavelets,[4] matching pursuit decomposition (MPD),[5] and statistical pattern recognition methods,[6] most of which involve a training-testing paradigm of classification. Essentially, these techniques rely on a training phase where a training dataset containing data from all classes is used to estimate parameters of statistical models that describe the classes. In the testing phase, test data is classified based on predictions from the learned class models. This procedure assumes that sufficient training data is available from each of the classes to be able to capture the statistics of the classes in a meaningful manner.[5,7] If the amount of data available for training is statistically insufficient, the performance of the resulting classification system can be poor due to inaccurate model parameters.

In real-world SHM applications, the collection of large amounts of training data is often difficult and can be time-consuming and resource intensive. However, a modest amount of data may be available from a few different but related experiments. This motivates the machine learning approach known as transfer learning (TL)[8] which is designed to provide a mechanism for information transfer between related learning tasks. In particular, TL can be a very effective tool for the efficient management and sharing of information between related classification tasks, and can help to reduce the training burden without significantly compromising on the classifier performance.[8]

Consider a learning scenario in which data is obtained from two experiments that are related. In the language of TL, one is referred to as the *source domain* experiment and the other as the *target domain* experiment. Each

Send correspondence to Debejyo Chakraborty, E-mail: debejyo@gmail.com, Web: http://www.debejyo.com
*Formerly with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ

Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2011, edited by Masayoshi Tomizuka, Chung-Bang Yun, Victor Giurgiutiu, Jerome P. Lynch, Proc. of SPIE Vol. 7981, 798147
 · doi: 10.1117/12.882025

domain is comprised of a task and a dataset. In the present problem, the learning task is that of classification and the dataset is comprised of the training and test data. TL intuitively assumes a natural flow of information from the source to the target domain (indeed, from a mathematical perspective, information can flow from source to target domain and vice versa).

In learning tasks such as classification, data for which the associated class is known is said to be *labeled*. Labeled training data then consists of training examples and their corresponding labels. Depending on the availability of labeled data (e.g., from the source or target domain), various types of TL are defined. In multi-task inductive transfer learning (MITL),[9–12] the source and target domains contain labeled data and the learning task is geared toward simultaneously improving the classification in both domains. In self-taught inductive transfer learning (SITL),[13] the source domain data is not labeled. In translated inductive transfer learning (TITL),[14] there is labeled data in the source domain and statistically insufficient labeled data in the target domain. These frameworks can be used to solve regression, classification, and cross-domain learning tasks. In transductive transfer learning (TTL),[15] only source domain labeled data is available. TTL is generally used to solve tasks that involve domain adaptation and covariate shift, regression and classification. In unsupervised transfer learning (UTL)[16–18] no labeled data is available. UTL is used to solve clustering and dimensionality reduction problems.

In this paper, a new TITL based structural damage classification method is proposed that makes use of statistics from related classification tasks to improve the classification performance on a data set with limited training examples. We demonstrate the utility of the proposed method for the classification of fatigue damage in an aluminum lug joint. Data was collected using four sensors, but sufficient training data was available only from one sensor which was considered as the source domain. The remaining three sensors with limited amount of data serve as the target domain. We confirm the penalty of using inadequate amount of training data for fatigue damage classification, and use the TITL based structural damage classification method to provide an effective solution. Our results show the dramatic improvement in damage classification performance when TITL is used to share information between the source and target domain sensors.

The remainder of this paper is organized as follows. Section 2 provides a background of the TL framework and describes the TL based classification approach. In Section 3, results are presented from an application of the proposed TL based classification method for classifying fatigue damage conditions in an aluminum lug joint sample. This is followed by concluding remarks in Section 4.

2. TRANSFER LEARNING FRAMEWORK

In this section, we briefly describe the analytical framework of transfer learning (TL) and Bayesian analysis utilized in the proposed classification method. For more details on these topics the reader is referred to the literature.[8,19]

2.1 Background on Classification

Given training data $\mathcal{Y}^{\mathrm{tr}} = \{Y_1^{\mathrm{tr}}, \ldots, Y_{N_{\mathrm{tr}}}^{\mathrm{tr}}\}$ with corresponding labels (class memberships) $\mathbf{c}^{\mathrm{tr}} = \{c_1^{\mathrm{tr}}, \ldots, c_{N_{\mathrm{tr}}}^{\mathrm{tr}}\}$, the conventional classification task entails learning a classifier $f : \mathscr{Y} \mapsto \mathscr{C}$ that can be used to classify test data $\mathcal{Y}^{\mathrm{te}} = \{Y_1^{\mathrm{te}}, \ldots, Y_{N_{\mathrm{te}}}^{\mathrm{te}}\}$. Here $\mathscr{Y}$ denotes the data space and $\mathscr{C}$ the class space. The class memberships are discrete and $\mathscr{C} = \{1, \ldots, M\}$ (for M possible classes).

Typically, the relationship between the data Y and class c is defined using probabilistic models of the form $P(Y|\boldsymbol{\theta})$ and $P(\boldsymbol{\theta}|c)$, with $\boldsymbol{\theta}$ being model parameters related to class c. In the Bayesian framework, a test data Y^{te} is classified as

$$c^{\mathrm{te}} = \underset{c \in \mathscr{C}}{\operatorname{argmax}}\ P\left(Y^{\mathrm{te}}|c\right)\ \Pr(c), \tag{1}$$

where the likelihood or predictive function $P(Y^{\mathrm{te}}|c)$ is given by

$$P\left(Y^{\mathrm{te}}|c\right) = \int P\left(Y^{\mathrm{te}}|\boldsymbol{\theta}\right)\ P(\boldsymbol{\theta}|c)\ d\boldsymbol{\theta}, \tag{2}$$

and $\Pr(c)$ is the *a priori* probability of the class c.

It is common practice to simplify the integral in (2) using a point approximation

$$P(\boldsymbol{\theta}|c) \approx \delta(\boldsymbol{\theta}, \boldsymbol{\theta}_c) \tag{3}$$

(δ is the Dirac delta function) so that (2) reduces to

$$P\left(Y^{\mathrm{te}}|c\right) = P\left(Y^{\mathrm{te}}|\boldsymbol{\theta}_c\right). \tag{4}$$

The parameters $\boldsymbol{\theta}_c$ can be estimated (in the 'training' step) using maximum-likelihood (ML) learning[19] on the subsets $\mathcal{Y}_c^{\mathrm{tr}} \triangleq \{Y_i^{\mathrm{tr}} \in \mathcal{Y}^{\mathrm{tr}} : c_i^{\mathrm{tr}} = c\}$ of the training data $\mathcal{Y}^{\mathrm{tr}}$ as

$$\boldsymbol{\theta}_c = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\ P\left(\mathcal{Y}_c^{\mathrm{tr}}|\boldsymbol{\theta}\right). \tag{5}$$

It is assumed that, for every class $c \in \mathscr{C}$, the number of data points in the training set $\mathcal{Y}_c^{\mathrm{tr}}$ is large enough so that the parameters $\boldsymbol{\theta}_c$ can be reliably estimated.

2.2 Transfer Learning for Classification

In the problem considered in this paper, the task is that of classification and we are specifically interested in the scenario where the amount of training data is small, i.e., the set $\mathcal{Y}^{\mathrm{tr}}$ is not of statistically significant size. As such, the parameters $\boldsymbol{\theta}$ required in the classifier (see Section 2.1) cannot be reliably estimated using $\mathcal{Y}^{\mathrm{tr}}$ and $\mathbf{c}^{\mathrm{tr}}$, leading to poor classification performance. However, when prior knowledge is available from a related task, it is possible to improve performance by learning a target domain predictive function that enables *transferring* of information between tasks using the framework of TL.[8] TL can be formally defined as follows: given source and target domain learning tasks, the goal of TL is to use the knowledge obtained from the source domain learning task to help improve the learning of the target domain predictive function.

In this paper, we describe a TITL based approach (simply referred to as TL in the remaining document) for improving classification performance in the target domain using prior classification information from the source domain. Specifically, both the source and target domain tasks are those of classification. Let $\mathcal{Y}_{\mathrm{sor}}^{\mathrm{tr}} = \{Y_1^{\mathrm{tr,sor}}, \ldots, Y_{N_{\mathrm{tr,sor}}}^{\mathrm{tr,sor}}\}$ be the source domain training data with corresponding class memberships $\mathbf{c}_{\mathrm{sor}}^{\mathrm{tr}} = \{c_1^{\mathrm{tr,sor}}, \ldots, c_{N_{\mathrm{tr,sor}}}^{\mathrm{tr,sor}}\}$ and $\mathcal{Y}_{\mathrm{sor}}^{\mathrm{te}} = \{Y_1^{\mathrm{te,sor}}, \ldots, Y_{N_{\mathrm{te,sor}}}^{\mathrm{te,sor}}\}$ the source domain test data. Similarly, $\mathcal{Y}_{\mathrm{tar}}^{\mathrm{tr}} = \{Y_1^{\mathrm{tr,tar}}, \ldots, Y_{N_{\mathrm{tr,tar}}}^{\mathrm{tr,tar}}\}$ is the target domain training data with class memberships $\mathbf{c}_{\mathrm{tar}}^{\mathrm{tr}} = \{c_1^{\mathrm{tr,tar}}, \ldots, c_{N_{\mathrm{tr,tar}}}^{\mathrm{tr,tar}}\}$ and $\mathcal{Y}_{\mathrm{tar}}^{\mathrm{te}} = \{Y_1^{\mathrm{te,tar}}, \ldots, Y_{N_{\mathrm{te,tar}}}^{\mathrm{te,tar}}\}$ the target domain test data. The size $N_{\mathrm{tr,tar}}$ of the target domain training data set $\mathcal{Y}_{\mathrm{tar}}^{\mathrm{tr}}$ is statistically insignificant. Our approach learns a target predictive function by relating the estimates of the parameters $\boldsymbol{\theta}^{\mathrm{sor}}$ from the source domain to the classification task in target domain, and improves the performance for classifying the target domain test data $\mathcal{Y}_{\mathrm{tar}}^{\mathrm{te}}$.

2.3 Formulation of Transfer Learning based Classification Approach

The training data in the source domain is used to probabilistically model M classes from which data is available. Under the assumption of point estimates for the model parameters computed using ML learning, the parameters $\boldsymbol{\theta}_c^{\mathrm{sor}}$ for the classes $c = 1, \ldots, M$ are estimated as

$$\boldsymbol{\theta}_c^{\mathrm{sor}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, P\left(\mathcal{Y}_{\mathrm{sor},c}^{\mathrm{tr}}|\boldsymbol{\theta}\right), \tag{6}$$

where $\mathcal{Y}_{\mathrm{sor},c}^{\mathrm{tr}} \triangleq \{Y_i^{\mathrm{tr,sor}} \in \mathcal{Y}_{\mathrm{sor}}^{\mathrm{tr}} : c_i^{\mathrm{tr,sor}} = c\}$ is the subset of the source domain training data $\mathcal{Y}_{\mathrm{sor}}^{\mathrm{tr}}$ belonging to class c.

Subsequently, the class memberships of the source domain test data are determined by performing classification:

$$c_j^{\mathrm{te,sor}} = \underset{c}{\operatorname{argmax}}\, P\left(Y_j^{\mathrm{te,sor}}|\boldsymbol{\theta}_c^{\mathrm{sor}}\right), \ \ j = 1, \ldots, N_{\mathrm{te,sor}}, \tag{7}$$

where the *a priori* probability of the classes has been assumed equal. The classification performance on test data validates how well the source domain model parameters have been estimated. Ideally, the amount of misclassification should be small.

In the target domain, the training and classification can be performed similarly using

$$\boldsymbol{\theta}_c^{\mathrm{tar}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, P\left(\mathcal{Y}_{\mathrm{tar},c}^{\mathrm{tr}} | \boldsymbol{\theta}\right), \tag{8a}$$

$$c_l^{\mathrm{te,tar}} = \underset{c}{\operatorname{argmax}}\, P\left(Y_l^{\mathrm{te,tar}} | \boldsymbol{\theta}_c^{\mathrm{tar}}\right), \quad l = 1, \ldots, N_{\mathrm{te,tar}}, \tag{8b}$$

with $\mathcal{Y}_{\mathrm{tar},c}^{\mathrm{tr}} \triangleq \{Y_k^{\mathrm{tr,tar}} \in \mathcal{Y}_{\mathrm{tar}}^{\mathrm{tr}} : c_k^{\mathrm{tr,tar}} = c\}$ the subset of the target domain training data $\mathcal{Y}_{\mathrm{tar}}^{\mathrm{tr}}$ belonging to class c. However, the training step in (8a) is likely to be inaccurate due to the statistically insufficient amount of training data $\mathcal{Y}_{\mathrm{tar}}^{\mathrm{tr}}$. This results in large misclassification when classification is carried out according to (8b).

In TL based classification, the target domain classification performance is improved by jointly using the information learned in the source domain and the data available in the target domain. Specifically, we modify the objective function to $P\left(Y_l^{\mathrm{te,tar}}, \boldsymbol{\theta}_c^{\mathrm{tar}}, Y_j^{\mathrm{te,sor}}, \boldsymbol{\theta}_c^{\mathrm{sor}}\right)$ and simplify as follows:

$$\begin{aligned} P\left(Y_l^{\mathrm{te,tar}}, \boldsymbol{\theta}_c^{\mathrm{tar}}, Y_j^{\mathrm{te,sor}}, \boldsymbol{\theta}_c^{\mathrm{sor}}\right) &= P\left(Y_l^{\mathrm{te,tar}} | \boldsymbol{\theta}_c^{\mathrm{tar}}, Y_j^{\mathrm{te,sor}}, \boldsymbol{\theta}_c^{\mathrm{sor}}\right) P\left(\boldsymbol{\theta}_c^{\mathrm{tar}}, Y_j^{\mathrm{te,sor}}, \boldsymbol{\theta}_c^{\mathrm{sor}}\right) \\ &\propto P\left(Y_l^{\mathrm{te,tar}} | \boldsymbol{\theta}_c^{\mathrm{tar}}, Y_j^{\mathrm{te,sor}}, \boldsymbol{\theta}_c^{\mathrm{sor}}\right) P\left(Y_j^{\mathrm{te,sor}} | \boldsymbol{\theta}_c^{\mathrm{sor}}\right), \end{aligned}$$

since $Y_j^{\mathrm{te,sor}}$ does no depend on $\boldsymbol{\theta}_c^{\mathrm{tar}}$. Thus, the classification in the target domain can be reformulated as

$$c_l^{\mathrm{te,tar}} = \underset{c}{\operatorname{argmax}}\, P\left(Y_l^{\mathrm{te,tar}} | \boldsymbol{\theta}_c^{\mathrm{tar}}, Y_j^{\mathrm{te,sor}}, \boldsymbol{\theta}_c^{\mathrm{sor}}\right) P\left(Y_j^{\mathrm{te,sor}} | \boldsymbol{\theta}_c^{\mathrm{sor}}\right). \tag{9}$$

In particular, let us consider the scenario where the source and target domain data are obtained from two different sensors, respectively. Under the assumption of conditional independence of the signals given class c, and using a typical additive independent random noise model (for example, independent and identically distributed white Gaussian noise), the measured data $Y_j^{\mathrm{te,sor}}$ and $Y_l^{\mathrm{te,tar}}$ are independent given the parameters $\boldsymbol{\theta}_c^{\mathrm{sor}}$ and $\boldsymbol{\theta}_c^{\mathrm{tar}}$. If, in addition, $Y_l^{\mathrm{te,tar}}$ does not depend on the source domain parameters $\boldsymbol{\theta}_c^{\mathrm{sor}}$, then (9) simplifies to

$$c_l^{\mathrm{te,tar}} = \underset{c}{\operatorname{argmax}}\, P\left(Y_l^{\mathrm{te,tar}} | \boldsymbol{\theta}_c^{\mathrm{tar}}\right) P\left(Y_j^{\mathrm{te,sor}} | \boldsymbol{\theta}_c^{\mathrm{sor}}\right). \tag{10}$$

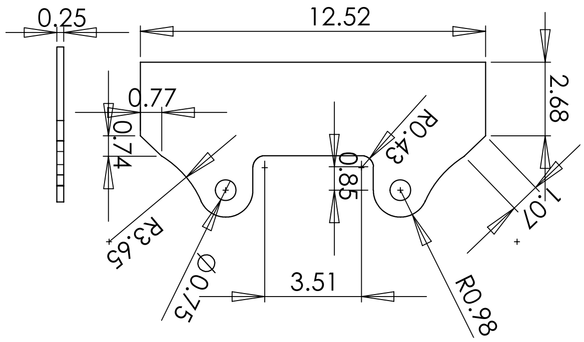
This TL based framework for information transfer and fusion of sensor data from the source domain is expected to provide superior classification performance in the target domain with limited training data.

3. APPLICATION TO STRUCTURAL DAMAGE CLASSIFICATION

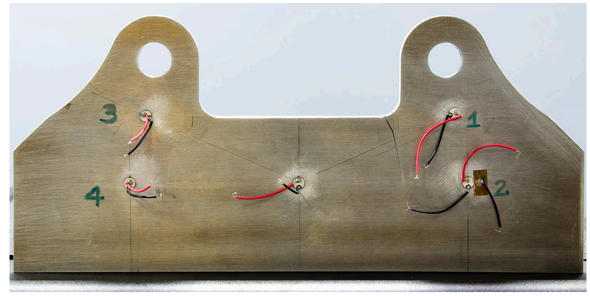
We now demonstrate the TL based classification method for classifying fatigue damage in an aluminum lug joint.

The aluminum lug sample considered in our experiments has dimensions as shown in Figure 1(a). Figure 1(b) shows the placement of the piezoelectric sensors used for data collection. There is one actuator and four sensors (numbered 1 through 4). Observe the presence of a crack close to sensors 1 and 2. The excitation signal is a 5-cycle tone-burst with center frequency 250 kHz, and is shown in Figure 2(a). In total there were $M = 11$ damage classes, corresponding to 2402, 22493, 32253, 62617, 72428, 82597, 112558, 154037, 154431, 155106 and 155705 cycles of fatigue loading (defining damage classes $c = 1, \ldots, 11$). For every damage class, 100 measurements were collected from each of the four sensors.

Matching pursuit decomposition (MPD)[20] with a Gaussian atom dictionary was used to extract time-frequency features[5] from the sensor signals. In particular, the MPD scaling parameter (see[5]) was restricted to be a function of the frequency, so that strictly 5 oscillations of the sinusoid were accommodated in a Gaussian window. This helped in reducing both the size of the dictionary and the dimensionality of the MPD features, and made the feature extraction process significantly faster (by a factor of 20). We considered $N = 10$ MPD iterations in this application. Figure 3 illustrates the MPD residual error as a function of the iteration number and the MPD based time-frequency representation (MPD-TFR)[20] for signals from two different damage classes.

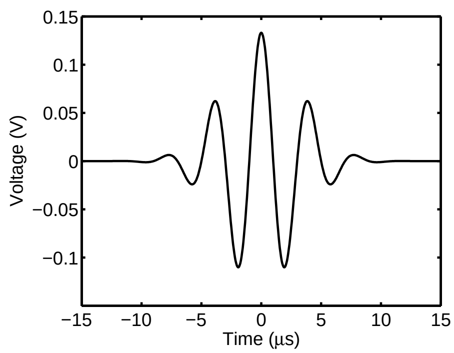


(a) Dimensions in inches.

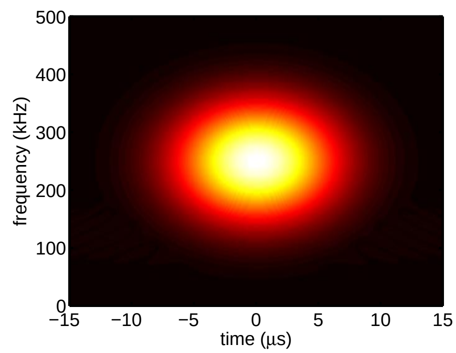


(b) Sensor placement and crack location.

Figure 1. Test sample geometry and sensor placement.



(a) Time-domain signal.



(b) Time-frequency representation.

Figure 2. Excitation signal.

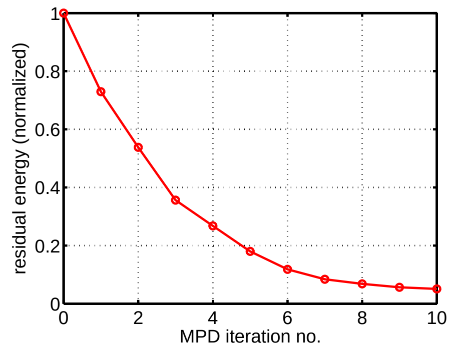
Let τ_n, ν_n, and α_n be the time-shift, frequency-shift, and coefficient in the nth term (iteration) of the MPD representation of a signal $s(t)$. The MPD representation can be viewed as a transformation of the signal $s(t)$ to a sequence:

$$s(t) \leftrightarrow \left[\begin{bmatrix} \tau_1^s \\ \nu_1^s \\ \alpha_1^s \end{bmatrix}, \begin{bmatrix} \tau_2^s \\ \nu_2^s \\ \alpha_2^s \end{bmatrix}, \ldots, \begin{bmatrix} \tau_N^s \\ \nu_N^s \\ \alpha_N^s \end{bmatrix} \right]$$

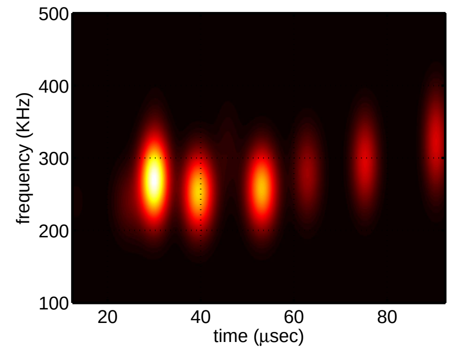
such that $\tau_1 \leq \tau_2 \leq \ldots \leq \tau_N$. This sequential representation is used to model signals from each damage class with a hidden Markov model (HMM).[21] Specifically, for each damage class, $\boldsymbol{\theta}_c = \{A_c, B_c, \boldsymbol{\pi}_c\}$, where A is the state-transition probability matrix, B is the observation density matrix, and $\boldsymbol{\pi}$ is the initial state distribution vector.[22] In our application, we utilized three-state HMMs for the models.

Sensors 1 and 2, which are closest to the damage, are expected to be most sensitive to the changes due to damage. Since sensor 1 was the sensor closest to the damage, it was selected as the source domain for TL. Sensors 2, 3 and 4 are used as target domains for TL. This is illustrated in Figure 3. In the source domain, 50 signals were used from each class for training. On the other hand, only 5 signals were used for training in each of the target domains. Training in the source and the target domains were conducted using Eqns (6) and (8a), respectively (see[22] for algorithmic details). Next, the source domain objective functions $P\left(\mathcal{Y}_{\mathrm{sor}}^{\mathrm{te}} | \boldsymbol{\theta}_c^{\mathrm{sor}}\right)$, $c = 1, \ldots, 11$, were evaluated and the class memberships $\mathbf{c}_{\mathrm{sor}}^{\mathrm{te}}$ were computed using (7). The corresponding test data from the target domains were classified both without using TL (according to (8b)) and using TL (Eqn (9)).

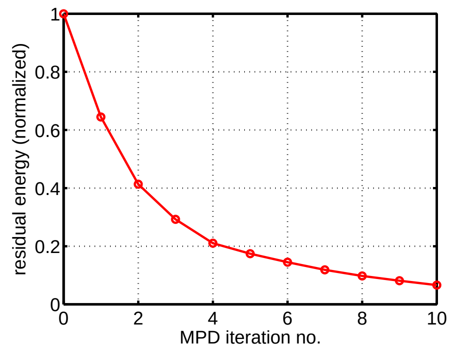
The damage classification results with and without using TL are shown in Figure 5. These figures demonstrate a confusion matrix which has 11 rows and 11 columns, each corresponding to a damage class. The rows represented the true damage classes and the columns represent the assigned damage classes. The colors correspond to the fraction of data that was correctly classified. For example, if 10% of the signals from class 1 are classified (incorrectly) to class 2, the entry at the intersection of row 1 and column 2 would have a value of



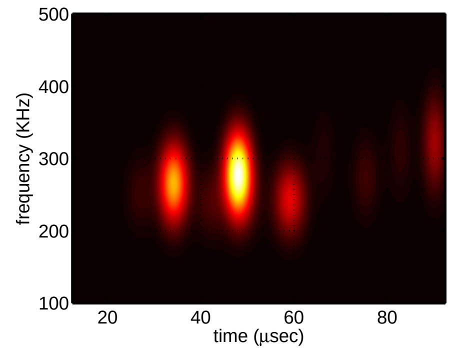
(a) MPD residual error for a signal from 2402 fatigue cycles.



(b) MPD-TFR of signal from 2402 fatigue cycles.



(c) MPD residual error for a signal from 155705 fatigue cycles.



(d) MPD-TFR of signal from 155705 fatigue cycles.

Figure 3. MPD residual error and MPD-TFR plots for a signal from sensor 1, corresponding to two different fatigue levels.

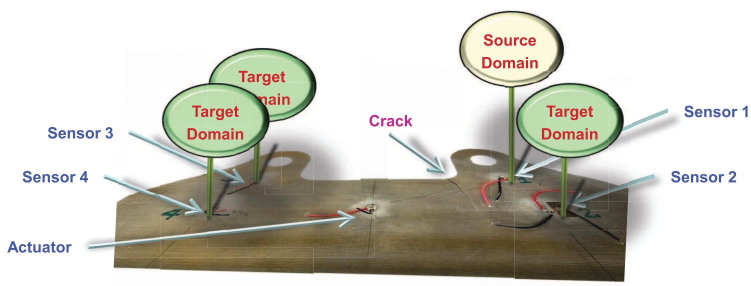
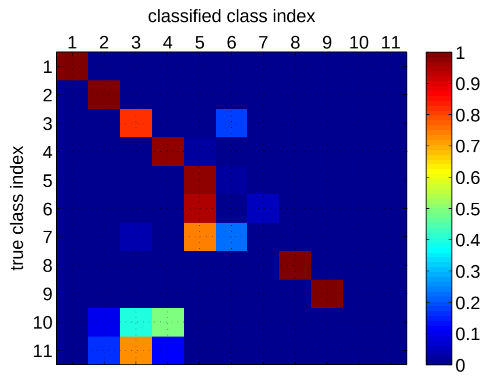


Figure 4. Transfer Learning domains used for fatigue damage classification.

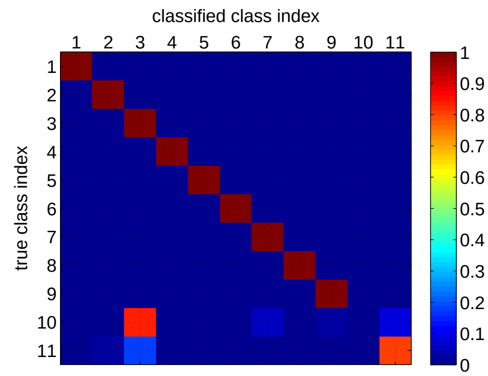
0.1. As we can see from the plots, using TL results in a significant improvement in the damage classification performance. This improvement is summarized in Table 1, which shows a comparison of the average correct classification rates before and after using TL.

Sensors	2	3	4
without TL	62%	2%	15%
with TL	89%	76%	65%

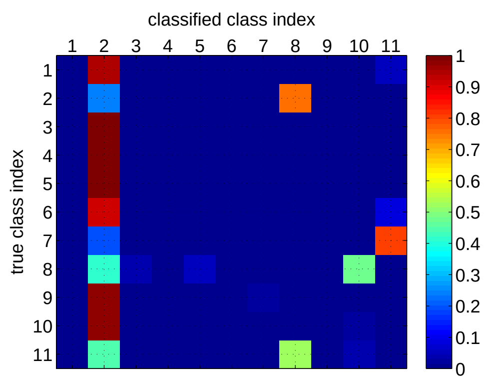
Table 1. Summary of improvement in damage classification performance using TL.



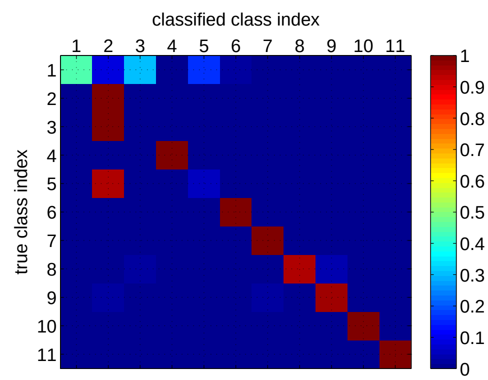
(a) Classification of sensor 2 signals (without TL).



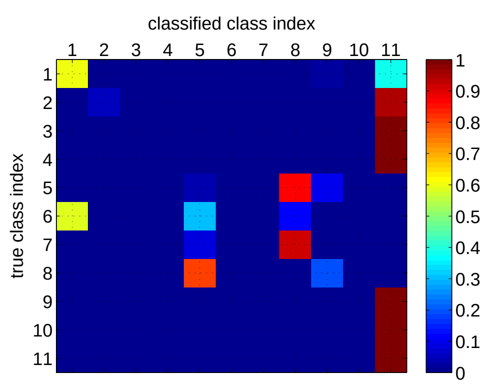
(b) Classification of sensor 2 signals (with TL).



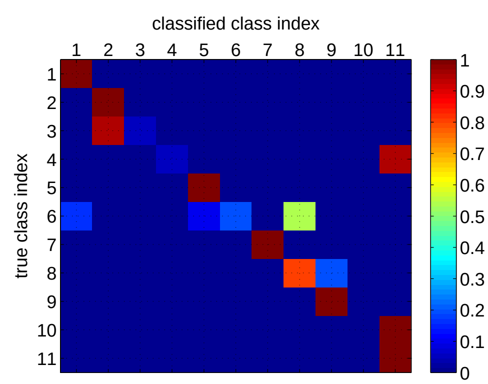
(c) Classification of sensor 3 signals (without TL).



(d) Classification of sensor 3 signals (with TL).



(e) Classification of sensor 4 signals (without TL).



(f) Classification of sensor 4 signals (with TL).

Figure 5. Graphical representation of confusion matrix showing damage classification results.

4. DISCUSSION

In this paper, we have demonstrated a novel transfer learning based structural damage classification method. Time-frequency based features were first extracted from the sensor signals using the MPD technique. The TF features were next modeled using HMMs, with a separate HMM used for data from each damage class. The TL based classification method then utilized the information that was learned from one sensor to enhance the classification performance on data obtained from another sensor. Our results show good performance of the method in classifying fatigue crack damage in a lug joint sample using very little training data.

In the particular SHM application considered here, the TL based classification is equivalent to Bayesian sensor data fusion. However, the TL based structural damage classification framework is more generally applicable in any setting where limited data available from several different but related experiments needs to be leveraged for best results.

ACKNOWLEDGMENTS

This research was supported by the MURI Program, Air Force Office of Scientific Research, grant number: FA9550-06-1-0309; Technical Monitor, Dr. David Stargel.

REFERENCES

[1] Farrar, C. R. and Worden, K., "An introduction to structural health monitoring," *Philosophical Transactions of the Royal Society A* **365**, 303–315 (2007).

[2] Staszewski, W. J., Boller, C., and Tomlinson, G., [*Structural Health Monitoring of Aerospace Structures, Smart Sensor Technologies and Signal Processing*], John Wiley and Sons, Ltd (2004).

[3] Yolken, H. T. and Matzkanin, G. A., "Recent trends in structural health monitoring technologies," *The AMMTIAC Quarterly* **3**(4), 3–6 (2008).

[4] Paget, C. A., Grondel, S., Levin, K., and Delebarre, C., "Damage assessment in composites by Lamb waves and wavelet coefficients," *Smart Materials and Structures* **12**(3), 393–402 (2003).

[5] Chakraborty, D., Kovvali, N., Wei, J., Papandreou-Suppappola, A., Cochran, D., and Chattopadhyay, A., "Damage Classification Structural Health Monitoring in Bolted Structures Using Time-frequency Techniques," *Journal of Intelligent Material Systems and Structures, special issue on Structural Health Monitoring* **20**, 1289–1305 (July 2009).

[6] Sohn, H., Farrar, C. R., Hunter, N. F., and Worden, K., "Structural health monitoring using statistical pattern recognition techniques," *Transactions of the ASME* **123**, 706–711 (2001).

[7] Chakraborty, D., Soni, S., Wei, J., Kovvali, N., Papandreou-Suppappola, A., Cochran, D., and Chattopadhyay, A., "Physics based modeling for time-frequency damage classification," in [*Proc. of SPIE, Smart Structures and Materials & Non-destructive Evaluation and Health Monitoring,*], **6926**, 69260M1–12 (2008).

[8] Pan, S. and Yang, Q., "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering* **22**, 1345 – 1359 (October 2010).

[9] Thrun, S., "Is learning the n-th thing any easier than learning the first?," in [*Advances in Neural Information Processing Systems*], 640–646, The MIT Press (1996).

[10] Caruana, R., "Multitask learning," *Machine Learning Journal* **28**(1), 41–75 (1997).

[11] Thrun, S. and Pratt, L. Y., eds., [*Learning to learn*], Kluwer Academic Publishers (1998).

[12] Baxter, J., "A model of inductive bias learning," *Journal of Artificial Intelligence Research* **12**, 149–198 (2000).

[13] Raina, R., Battle, A., Lee, H., Packer, B., and Ng, A. Y., "Self-taught learning: Transfer learning from unlabeled data," in [*Proceedings of the 24th international conference on Machine learning*], (2007).

[14] Dai, W., Chen, Y., Xue, G.-R., Yang, Q., and Yu, Y., "Translated learning: Transfer learning across different feature spaces," in [*Proceedings of the 22nd Annual Conference on Neural Information Processing Systems (NIPS-08)*], (2008).

[15] Arnold, A., Nallapati, R., and Cohen, W. W., "A comparative study of methods for transductive transfer learning," in [*7th IEEE International Conference on Data Mining Workshops*], 77 – 82 (2007).

[16] Hinton, G. and Sejnowski, T. J., eds., [*Unsupervised Learning: Foundations of Neural Computation*], MIT Press (1999).

[17] Duda, R. O., Hart, P. E., and Stork, D. G., [*Pattern classification*], ch. Unsupervised Learning and Clustering, 571, Wiley, New York, 2 ed. (2001).

[18] Kotsiantis, S. and Pintelas, P., "Recent advances in clustering: A brief survey," *WSEAS Transactions on Information Science and Applications* **1**(1), 73–81 (2004).

[19] MacKay, D. J. C., [*Information Theory, Inference, and Learning Algorithms*], Cambridge University Press (2003).

[20] Mallat, S. G. and Zhang, Z., "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal Processing* **41**, 3397–3415 (December 1993).

[21] Rabiner, L. R., "A tutorial on hidden Markov models and selected applications in speech recognition," in [*Proceedings of the IEEE*], **77**, 257–286 (1989).

[22] Zhou, W., Chakraborty, D., Kowali, N., Papandreou-Suppappola, A., Cochran, D., and Chattopadhyay, A., "Damage classification for structural health monitoring using time-frequency feature extraction and continuous hidden Markov models," in [*Asilomar Conference on Signals, Systems and Computers*], 848–852 (November 2007).