A hierarchical classification scheme for computationally efficient damage classification

C K Coelho[1]*, **S Das**[2], and **A Chattopadhyay**[1]
[1]Department of Mechanical and Aerospace Engineering, Arizona State University, Tempe, Arizona, USA
[2]UARC, UCSC, NASA Ames Research Center, Moffett Field, California, USA

The manuscript was received on 23 July 2008 and was accepted after revision for publication on 6 May 2009.

DOI: 10.1243/09544100JAERO428

Abstract: This article presents a methodology for data mining of sensor signals in a structural health monitoring (SHM) framework for damage classification using a machine-learning-based approach called support vector machines (SVMs). A hierarchical decision tree structure is constructed for damage classification and experiments were conducted on metallic and composite test specimens with surface mounted piezoelectric transducers. Damage was induced in the specimens by fatigue, impact, and tensile loading; in addition, specimens with seeded delaminations were also considered. Data were collected from the surface mounted sensors at different severities of induced damage. A matching pursuit decomposition (MPD) algorithm was used as a feature extraction technique to preprocess the sensor data and extract the input vectors used in classification. Using this binary tree framework, the computational intensity of each successive classifier is reduced and the efficiency of the algorithm as a whole is increased. The results obtained using this classification show that this type of architecture works well for large data sets because a reduced number of comparisons are required. Due to the hierarchical set-up of the classifiers, performance of the classifier as a whole is heavily dependent on the performance of the classifier at higher levels in the classification tree.

Keywords: damage classification, support vector machines, matching pursuit decomposition, hierarchical decision tree, structural health monitoring

1 INTRODUCTION

Aerospace components are often subjected to a variety of mechanical, thermal, and environmental loading conditions that are not considered during the design of the part. This 'over-matching' in the design highlights the need for condition-based maintenance in place of the current schedule-based maintenance practice. A robust system for condition-based health management relies heavily on an accurate method for damage assessment so that prognostic decisions can be made. The methods currently used for damage detection, such as acoustic emission [**1**], thermography [**2**], eddy current method [**3**], and ultrasonic scanning [**4**], require that the structure be taken out of service and sometimes disassembled for inspection. The weight and cost of these systems make it very impractical for on-board applications. Some of the most promising candidates for *in-situ* damage detection are piezoelectric transducers [**5**] and fibre optic sensors [**6**]. In this article, piezoelectric transducers are used because they are light weight, cost effective and can be used for both passive and active damage detection.

Damage detection involves the characterization and identification of the key damage-related features, which can be used within a structural health monitoring (SHM) framework as a damage indicator. Physics-based modelling techniques can be used to accurately characterize the interaction of the actuation-induced stress wave with a given damage and sample type to determine the sensor output [**5**]. The drawback of this approach is that it is very computationally expensive to model all the expected damage scenarios for a particular part and even then it does not account for imperfections in the material and sensors, unexpected damage, and

*Corresponding author: Department of Mechanical and Aerospace Engineering, Arizona State University, P.O. Box 876106, Tempe, Arizona 85287-6106, USA.
email: clyde.coelho@asu.edu

ambient noise. Also, it is very difficult to solve the inverse problem of finding the change in the system due to damage using the sensor signal alone. The methodology used for this research is a data-driven approach [**7**] that uses examples of actual signals from sensors that have been categorized by experts and uses them as a guide to identifying similar damage types. It is assumed that the training data collected from every class of damage contains information within the signals that relates to the wave–damage interaction. Although this method requires a large database of training signals, it takes into account all the variations that are not accounted for in physics-based models.

In this article, support vector machines (SVMs), an advanced machine-learning-based technique, have been utilized. SVMs have a strong mathematical foundation [**8**] and show good generalization when applied to classification problems in a number of fields. Traditional approaches to the classification of multiclass problems have been conducted in the form of 'one versus one', 'one versus rest', hybrid [**9**, **10**], and clustering [**11**] algorithms. In 'one versus one', the amount of training time required is very large since $k(k-1)/2$ classifiers need to be constructed for a k class problem. For a 'one versus rest' scheme, a problem involving k classes of data will require the construction of k classifiers. One problem with the latter case is that each classifier will require the use of the entire training set, which becomes computationally intractable. For both methods, a voting scheme is used in which the classifier that scores the highest for a given data set assigns all the points in that particular set to a given class. Also, in such a case, it is very difficult to decide which class the test data belongs to if two classifiers have similar scores. Clustering schemes are able to learn signal characteristics well and can decide the uniqueness of different classes (or even classes within classes) based on the clustering of data points in a hyperspace. While this approach is promising for damage detection scenarios where all possible damage types cannot be known, the computational expense involved with determining the cluster boundaries increases exponentially with the increase in training sets.

This article presents a framework for damage classification that applies a 'one versus rest' scheme [**12**] organized into a binary tree structure that addresses some of the computational issues associated with classification involving a large number of classes by reducing the total number of required classifiers while accurately classifying damage in composite and metallic structures. The test specimens were instrumented with surface mounted transducers and then subjected to different damage-inducing mechanisms. Data were collected at various damage states for analysis. A matching pursuit decomposition (MPD) [**13**] algorithm is used to extract the features of the collected signal prior to classification by the binary tree SVMs framework.

The novelty of this approach over traditional SVM schemes is twofold. First, it allows the user to prioritize damage cases, making it quicker to identify common or expected damage states. The other main advantage is that by organizing classifiers correctly, it is possible to simultaneously reduce the number of classifiers necessary as well as the complexity of each classifier. The research presented in this article shows that this organization works well for these types of applications.

This article is organized as follows. Sections 2 and 3 present a theoretical background on the method used for classification and its organization as a binary tree classifier. The theory behind the extraction of features that are used for classification is presented in section 4. The experimental set-up and some details regarding data collection have been discussed in section 5. Section 6 demonstrates the effectiveness of the classification scheme through some selected results from a fatigued lug joint. Section 7 presents some observations made from the present study along with suggestions for further improvement of the current work.

2 SUPPORT VECTOR MACHINES

SVMs have been used for classification in a number of different fields because of their good generalization [**12**] ability, which means they are able to learn the behaviour of the system even with relatively few examples. The ability of SVMs to separate non-linearly separable data is based on Cover's theorem [**14**], which states that non-separable or non-linearly separable patterns in input space (low dimensional space) are more likely to be linearly separable in a new high-dimensional feature space, provided that the transformation is non-linear and the dimensionality of the feature space is high enough. The patterns in this high-dimensional (say N dimension) space are then separated by constructing an $N-1$ dimension hyperplane, which takes into account the non-linear relationship of the data. The mapping kernel used in this research is the radial basis function (RBF), which is popular in machine learning applications [**15**, **16**] involving data sets that are not linearly separable. The RBF kernel takes the form

$$K(\boldsymbol{x}, \boldsymbol{x}_i) = \mathrm{e}^{-\|\boldsymbol{x}-\boldsymbol{x}_i\|/2\sigma^2} \tag{1}$$

where $\boldsymbol{x}$ is the input vector, $\boldsymbol{x}_i$ is the ith input pattern, and σ is the width of the kernel that has been optimized during the training phase. Once the training data have been mapped into high-dimensional space, an optimal hyperplane is constructed. The optimal hyperplane is defined as one which maximizes [**14**] the separation between the two classes, allowing the algorithm to learn the differences in the data sets.

The decision boundary for patterns that are linearly separable is defined as

$$\boldsymbol{w}z_i + b = 0 \tag{2}$$

where $\boldsymbol{w}$ is an adjustable weight vector, b is a bias, and z_i is the mapped input pattern. For an input point (z_i, y_i), where y_i is the corresponding class label, the above equation is subject to

$$y_i(\boldsymbol{w}z_i + b) \geqslant 1 \tag{3}$$

This is obtained by rescaling $\boldsymbol{w}$ and b to get two parallel hyperplanes for $y_i = \pm 1$. The distance between these two hyperplanes defines the margin of the classifier. In most practical applications, the data are non-linearly separable and it is not possible to construct a hyperplane without admissible training errors. In such a case, a soft margin [**17**] is imposed which allows a certain number of data points to be misclassified (Fig. 1) and the above equation can be modified as

$$y_i(\boldsymbol{w}z_i + b) \geqslant 1 - \xi_i \tag{4}$$

where $\xi_i \geqslant 0 \ \forall i$ is a slack variable. In order to find this optimal hyperplane, which minimizes classification error, the following optimization problem needs to be solved

$$\min \frac{1}{2}\|\boldsymbol{w}\|^2 + D\sum_i \xi_i \tag{5}$$

subject to the constraint shown in equation (4). The variable D refers to a regularization parameter that can be modified to control the complexity of the model. A large value of D means that the classifier will only classify separable data. For this research, a k-fold cross validation scheme will be used to optimize the hyperparameters σ and D. Defining $w(\alpha) = \sum_i \alpha_i y_i z_i$, the dual problem can be constructed as

$$\max W(\alpha) = \sum_i \alpha_i - \frac{1}{2}w(\alpha)w(\alpha) \tag{6}$$

subject to $0 \leqslant \alpha_i \leqslant D \ \forall i$ and $\sum_i \alpha_i y_i = 0$. Solving equation (6) for the Lagrange multipliers (α), it is possible to recover the solution to the primal problem. The decision function for the classifier becomes

$$y(x) = \text{sign}\left\{\sum_i \alpha_i y_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b\right\} \tag{7}$$

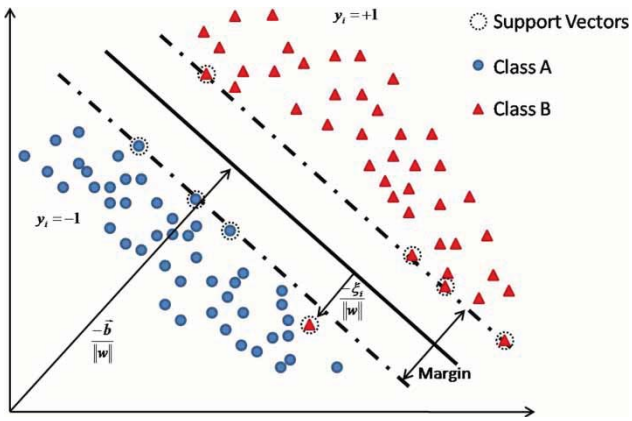

Fig. 1 Representation of parameters needed for hyperplane construction in two dimensions

3 SVMS AS A BINARY TREE CLASSIFIER

The multi-class classifier used in this article combines a modified 'one versus rest' algorithm with a binary tree structure [**12**] to minimize the number of comparisons that are necessary to identify a data class while still taking into account all possible classes. For a four-class problem as shown in Fig. 2, the binary tree classifier is set-up as follows.

1. A two-class SVM is trained using pattern 1 as Class A and patterns 2, 3, and 4 as Class B and a hyperplane is constructed.
2. Next, data points corresponding to pattern 1 are removed from the training set and pattern 2 is denoted as Class A and patterns 3 and 4 are denoted as Class B for hyperplane construction.
3. This process is repeated until the last classifier compares the last two patterns.

The advantage of this approach is that for a k-class problem, only $(k-1)$ hyperplanes need to be constructed. Also, removing patterns after each classifier is constructed reduces the computational expense. In this way, it is also possible to prioritize damage classes and terminate the classification algorithm during testing before checking all possible cases. In this article, since the damage considered is only of one type, the classifiers are arranged in order of increasing crack length. In a more complex structure like a bolted joint for example, it would be possible to prioritize torque loss (fault type A) over structural damage (fault type B) and the algorithm can be terminated mid way if a loose bolt is detected [**7**, **15**]. In order to ensure there

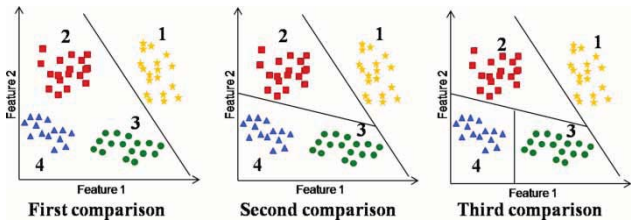

Fig. 2 Construction of multiple hyperplanes without overlapping regions for multi-class problems

is no region of overlap where one point could be classified as belonging to multiple classes, a point in the first comparison that is classified as belonging to Class A is removed from the test set and the points in Class B move on to the next classifier.

4 FEATURE EXTRACTION ALGORITHM

Feature extraction is a vital part of any data mining algorithm. The features extracted from the raw data need to be meaningful in the sense that they reflect changes in the system due to different types of damage. In this research, MPD has been used as a feature extraction tool. MPD has been used for various applications such as feature extraction, signal characterization and classification [**18**], and signal encoding and reconstruction [**19**]. The working principle of MPD relies on decomposing a given signal into linear expansions of elementary functions (or atoms). The resulting decomposition reveals the waveform's time–frequency structure [**13**]. A change in the signal is represented by the selection of different atoms that represent the waveform.

In this research, the dictionary elements were composed of Gabor atoms, normalized in both the time and frequency domain [**13**]. These atoms were selected since they have energy that is concentrated in the time–frequency domain and there exists a closed-form [**13**, **20**] analytical time–frequency representation for such atoms. Also, the algorithm is guaranteed to converge if the dictionary used is complete and the atoms have unitary energy [**21**]. The decomposition of the signal is based on four variables that define each dictionary element: expansion coefficient (C), time shift (τ), frequency shift (f), and atom width (k). The expression for the atoms used is given by

$$g_i(t) = \mathrm{e}^{-k_i^2(t-\tau_i)^2 \cos(2\pi f_i t)} \tag{8}$$

Using these atoms, the decomposition after M iterations can be written as

$$x(t) = \sum_{i=0}^{M-1} C_i g_i + R^M x(t) \tag{9}$$

where $R^M x(t)$ is the residue of the signal after decomposing the signal M times and $R^0 x(t)$ is the original signal for $M = 0$. As $M \to \infty$, the signal residue will approach zero and the entire signal will be decomposed, that is

$$\lim_{M\to\infty} \|R^M x(t)\|_2 = \lim_{M\to\infty} \|x(t) - \sum_{i=0}^{M-1} C_i g_i\|_2 = 0 \tag{10}$$

The MPD algorithm is adopted because it reduces a given signal into fewer representative components that are more easily classified. Also, for physical systems, the number of iterations can be limited so that the part of the signal that contains information is decomposed while the noise is contained in the residue. First, the weighted contribution of the dictionary element that best matches the signal (or the residue) is calculated. The dictionary element that has the highest time correlation with the signal is selected and the weighted element is then extracted from the signal. The signal residue that is left is put back into the algorithm until the stopping criteria is reached. The stopping criteria can be defined in terms of the minimum energy that is extracted from the signal or the total number of iterations of the algorithm.

5 EXPERIMENTAL SET-UP

For this study, experiments were carried out using different damage mechanisms in metallic and composite materials. The damage types investigated were fatigue cracking in aluminium lug joints along with delaminations, impact damage, and tensile failure in carbon fibre composite plates.

5.1 Fatigue crack

The specimen tested was a lug joint that was subjected to tensile fatigue loading as shown in Fig. 3(a). The sample was machined out of Al 2024 T351. One surface of the lug joint was polished using 1200 grit silicon carbide paper so that more accurate measurements of crack length could be made using an optical telescope. The sample was tested at a load ratio of 0.1 with a maximum load of 1100 lbs at 20 Hz using an Instron 1331 servohydraulic test frame. Images of the crack length were taken every time the test was halted for data collection from the piezoelectric transducers

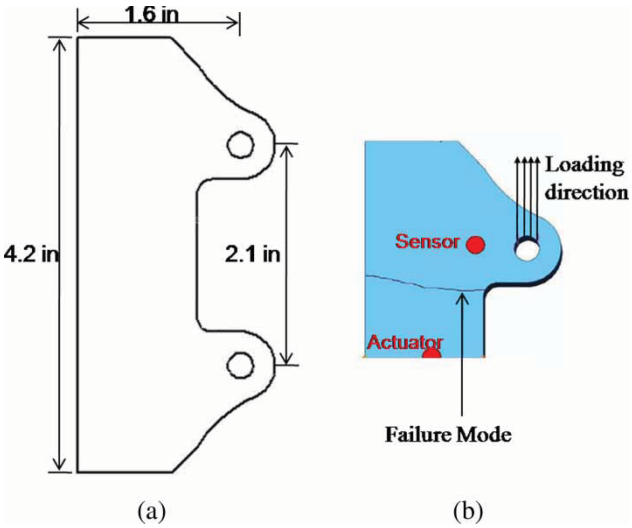

Fig. 3 (a) Specimen dimensions and (b) specimen with sensor/actuator placements and failure modes

using a CCD camera. Figure 3(b) shows the placement of the actuator and sensors on the structure. For the active interrogation and detection scheme used in this research, a 130 kHz, Gaussian windowed sine wave was used as the excitation signal. The duration of the excitation was 500 μs. The data collected from the sensors were sampled at 2 MHz. Before preprocessing, each observation was downsampled to 500 kHz with a signal length of 512 points. Downsampling was feasible since the excitation was a narrow band and most of the components of the sensor signal were between 100 and 150 kHz and the Nyquist frequency was still well above the maximum frequency component of the signal. It also made the matching pursuit algorithm more computationally efficient as the required dictionary size is reduced. A total of 300 observations were taken every time the damage state was measured. The fatigue experiment carried out on the lug joint resulted in five different damage states being measured. The damage states differed in the length of the crack that was present. The different damage states are represented by Ci where i represents a different damage class, corresponding to (a) C1: Healthy; (b) C2: 6 mm crack (27.1 per cent); (c) C3: 8 mm crack; (36.2 per cent); (d) C4: 10 mm crack (45.3 per cent); and (e) C5: 12 mm crack (54.3 per cent). The dimensions in parentheses are relative crack lengths with respect to a total possible crack length (width of the sample) of 1.15 in.

5.2 Delamination

Four 12″ × 12″ composite plates each with different delaminations cases were tested as shown in Fig. 4. The composite material used for all of the composite tests was a HexPly 954-3 unidirectional carbon fibre with a cyanate resin in a $[0, 90]_{4s}$ configuration. Again, a Gaussian windowed sine wave was used as the excitation and a total of 360 signals were measured for every different delamination scenario. A sampling frequency of 500 kHz was used when acquiring the data and only data from sensor 3 were used when classifying the data. The different delamination cases tested were (a) C1: Healthy; (b) C2: 5 per cent delamination at the fourth interface; (c) C3: Square delamination at the edge of the fourth interface; and (d) C4: Square delamination on the corner of the fourth interface.

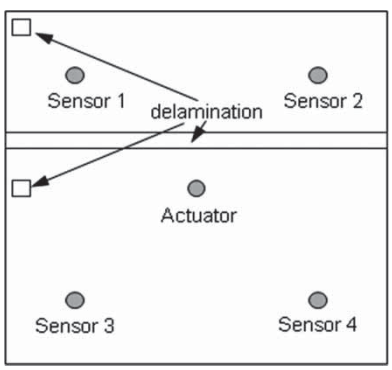

Fig. 4 Composite plate with delamination and sensors

5.3 Impact

The coupons used for the impact tests are shown in Fig. 5. Each coupon was impacted at a different velocity and sensor responses due to a Gaussian windowed sine wave excitation were collected. The sampling frequency used for signal acquisition was 2 MHz. A total of 150 signals were collected after impacting the sample. For the impact test performed, the different damage types were (a) C5: Healthy; (b) C6: Impact velocity of 2.53 m/s; (c) C7: Impact velocity of 2.11 m/s; and (d) C8: Impact velocities of 1.71 m/s.

5.4 Tensile damage

The tensile damage in the composite specimen shown in Fig. 6 was induced by first putting a notch in the

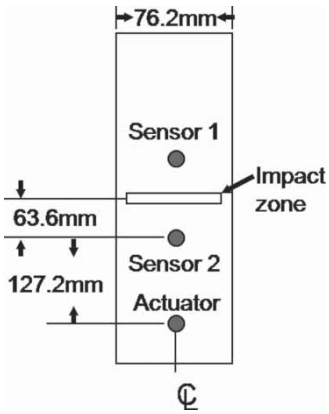

Fig. 5 Dimensions of impact specimen

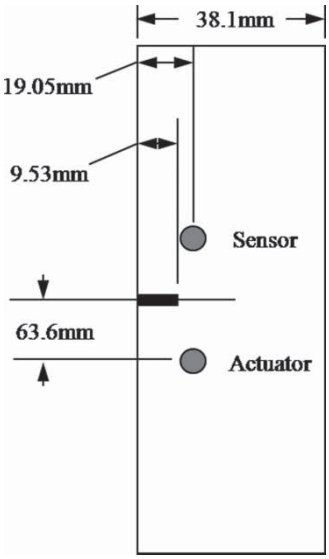

Fig. 6 Tensile test specimen

specimen so that a stress concentration is created and then loading the specimen until failure. The load levels where data was collected were determined by listening to the sample being loaded until a cracking sound was heard. Then the sample was unloaded and data were collected again using a Gaussian windowed sine wave for actuation. The different damage levels are (a) Class 9: Healthy; (b) Class 10: 5350 lb tensile loading; (c) Class 11: 6500 lb tensile loading; (d) Class 12: 7700 lb tensile loading; and (e) Class 13: 8800 lb tensile loading.

Table 1 Results of the nested binary classification scheme (lug joint)

	Predicted class				
Actual class	C1	C2	C3	C4	C5
C1	94	1	2	1	2
C2	6	88	2	1	3
C3	8	7	74	4	7
C4	3	3	5	85	4
C5	10	2	4	5	79

6 RESULTS AND DISCUSSION

Data collected from the experiments were first passed through the MPD algorithm to extract features that are more easily classifiable. Figure 7(a) shows the first three principal components extracted using principal component analysis (PCA) of the raw signals that were collected during the testing of the aluminium lug joint. It can be seen from this figure that there is a tremendous amount of overlap and that accurate separation of these points in this form will be extremely difficult. The true dimension of the data being analysed is 512. Figure 7(b) shows a PCA plot of the same signals after feature extraction. The feature extraction procedure reduced the dimension of the data from 512 to 60. It can be seen that even though there is still some overlap in the points, separation of these points in high-dimensional space is made easier.

When classifying the damage state of the lug joint, training of each classifier in the binary tree was completed using 200 examples that belonged to each class. The testing of the classifier was completed using 100 data points from each damage class. A five-fold cross validation was used to optimize the hyperparameters. The results of the classification algorithm are presented in the nested binary confusion matrix (Table 1). From Table 1, it can be seen that the classifier performs extremely well in identifying data points that belong to C1, especially considering that a 15 per cent error was permitted when training the classifier. This percentage was selected because of the nature of the overlap of the data patterns. It also prevented the classifier from 'over-fitting' the hyperplane to the data resulting in a loss of generalization.

For data points belonging to other classes, a small but significant portion of the data was misclassified as belonging to C1. Since points that are positively classified are removed from the test set before further classification, the elements in every column are classification results from a smaller set of data. As an example, the number of points in C2 that were correctly classified was 70 out 80 test points. A drawback of this classification scheme is that the results of a classifier are dependent on the performance of classifiers that are evaluated at higher nodes in the tree.

In order to better visualize the data overlap, a histogram of the normalized distance of all the points from the hyperplane is constructed (Fig. 8). The histogram clearly shows that when testing the decision hyperplane, there are relatively few points from C1 that are mistakenly classified but there is a much larger number of points belonging to C2–C5 that fall into the C1 side of the decision plane. This causes the relatively large misclassification of points belonging to C2–C5. An optimal selection of training points from C2 to C5 may allow for the construction of a more accurate classifier for this case. This will be investigated in future work. Chakraborty *et al.* [**22**] analysed the data from this experiment and generated classification results

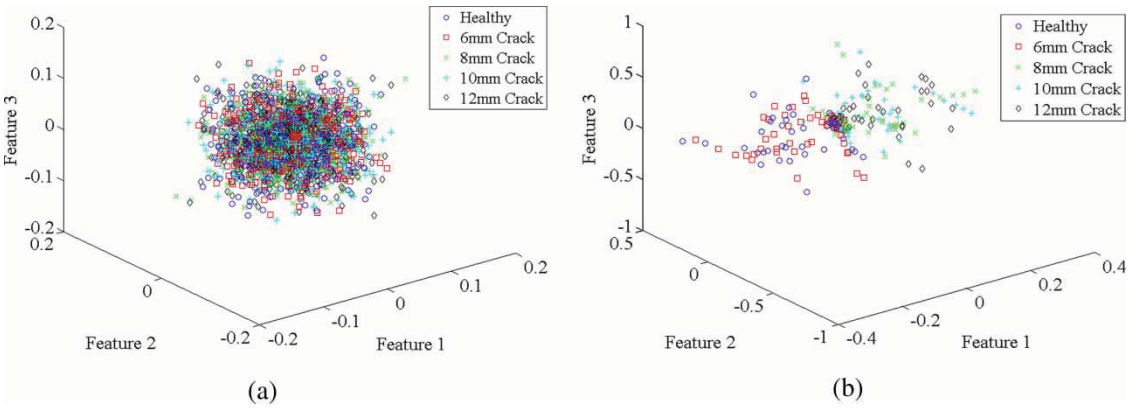

Fig. 7 PCA of (a) raw signals and (b) MPD features extracted from signals

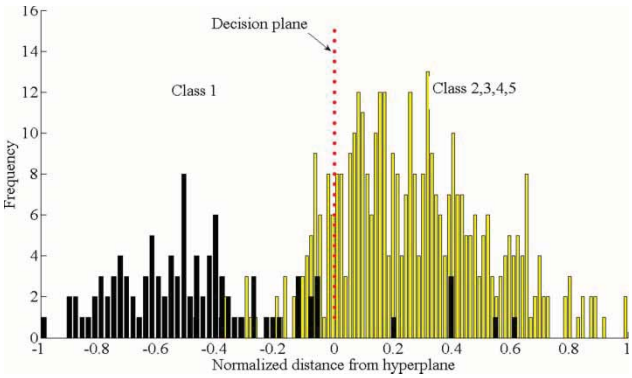

Fig. 8 Histogram of distance from the optimal hyperplane

using a Hidden Markov Model (HMM) with 20 MPD iterations. The results obtained have a minimum correct classification rate of 88.6 per cent as compared to the 84 per cent average classification presented for this work in Table 1. While resulting in a slightly lower classification than the HMM algorithm, the binary tree SVM framework proposed here takes substantially less time to run. This feature will become more prominent when the size of the data sets becomes large as in real world applications. A quantitative comparison of CPU run times and classification results will be carried out on larger data sets in future work.

Figure 9 shows the receiver operating characteristic (ROC) curve [**23**] for each of the constructed classifiers. An ROC curve is a robust way to test the ability of a classifier to discriminate between classes. It allows a user to weigh the cost savings from maintaining or replacing a part after it is damaged but just before failure (true positive) against the added cost of replacing a part when it is still healthy (false positive). An ideal classifier would have a point at (0, 1), which means that the classifier was able to correctly identify all the damage states and there was no overlap in the data patterns. If the data patterns do have some overlap, then false positives would occur, since the decision plane is fixed after training. The five classes being studied result in the construction of four classifiers

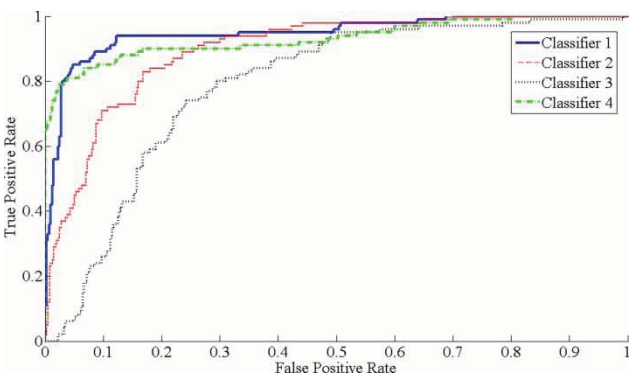

Fig. 9 ROC curve for each classifier used

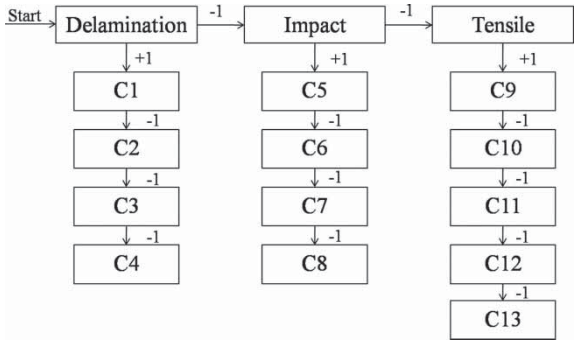

Fig. 10 Organization of binary tree classifier

that are constructed at different levels of the binary tree. The curves for each classifier represent the performance of each individual classifier and should not be used to judge the performance of the entire classification scheme. It can be seen from this plot that all the classifiers have a performance well above the line of no-discrimination.

Next the data from the composite experiments were analysed in one large classification framework as shown in Fig. 10. In the first level of the tree, the algorithm determines whether the damage class is delamination, impact, or tensile. Table 2 shows the results of the first level of the classifier. For this classification, two-thirds of the data was used for training and the remaining data was used for testing. It can be seen that the classifier is able to easily tell the difference between the different damage types because the interaction of the excitation signal with the different damage cases produces very different output signals.

Once the class of damage is selected, the algorithm then goes on to discard data from the other damage types and then determine the extent of the determined damage type. Tables 3 to 5, which represent level 2 in the hierarchy, show the results of the classifier within each of these damage types. In each of

Table 2 Classification of damage-type results (composite)

	Predicted class		
Actual class	Delamination	Impact	Tensile
Delamination	1	0	0
Impact	0	0.995	0.005
Tensile	0	0	1

Table 3 Classification results for delamination

	Predicted class			
Actual class	C1	C2	C3	C4
C1	0.88	0.02	0.04	0.06
C2	0.04	0.86	0.04	0.06
C3	0.06	0.04	0.90	0.00
C4	0.12	0.08	0.06	0.74

Table 4 Classification results for impact damage

	Predicted class			
Actual class	C5	C6	C7	C8
C5	0.88	0.04	0.00	0.08
C6	0.10	0.82	0.02	0.06
C7	0.06	0.02	0.90	0.02
C8	0.02	0.04	0.04	0.90

Table 5 Classification results for tensile damage

	Predicted class				
Actual class	C9	C10	C11	C12	C13
C9	0.90	0.04	0.02	0.02	0.02
C10	0.06	0.88	0.02	0.00	0.04
C11	0.08	0.04	0.80	0.06	0.02
C12	0.04	0.02	0.04	0.86	0.04
C13	0.08	0.00	0.04	0.04	0.84

the level 2 classifiers, 100 data points from each class were used for training and 50 points were used for testing. The classification rates of the level 2 classifiers are not as good as the level 1 classifiers because there is more overlap in the sensor signals of a particular damage type. In the case of the seeded delaminations (Table 3), it can be seen that the classifier performs very well when identifying C1–C3 but has a little trouble correctly classifying C4. This is probably because the damage is located very far from the sensor and all of the reflected waves from the corner delamination do not reach sensor 3. The high-energy components that do reach the sensor are common to the signals in C1–C3, causing them to be misclassified. In the case of damage cause by impact and a tensile test (Tables 4 and 5), the classification tool is able to accurately classify all the damage classes C5–C13, indicating that the constructed hyperplane is able to separate the classes while avoiding significant overlap.

Figure 11 shows the computational time required to build the classification tree using sensor data from the tensile test on a composite plate. It can be seen from Fig. 11 that in the case of a three-class problem, the performance of the 'one-versus-one' method and the binary tree method are similar but the 'one-versus-all' approach is not as efficient. This is because the overlap in the damage classes is significant and the time taken to optimally select the hyperparameters for a 'one versus all' approach is longer than the time requirement for extra comparisons required in the 'one-versus-one' approach. However, when the number of classes becomes larger, the number of comparisons increases and this starts to increase the execution time rapidly as shown in the case when five classes are considered. It is expected that if more data classes of this type were available, the 'one-versus-one' approach would be less efficient than the 'one-versus-all' approach. In the binary tree approach, the first comparison has the same computational intensity as a 'one versus all'

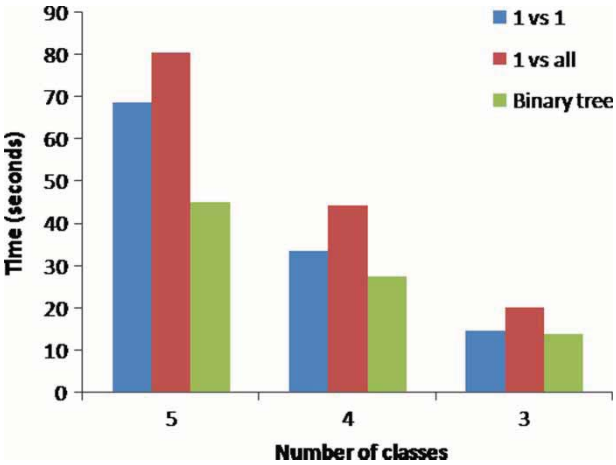

Fig. 11 Computational efficiency of different SVM approaches

approach. As the algorithm proceeds along the tree path, the complexity of each successive classifier is reduced until the last node where it has the same computational intensity as a 'one versus one' classifier. This feature, combined with the reduced number of comparisons necessary for a k-class problem, makes it more efficient, as shown in Fig. 11.

7 CONCLUDING REMARKS

A framework has been developed incorporating SVM classifiers organized into a binary tree structure with an MPD algorithm that is used for feature extraction. The goal was to classify different damage states in structural hotspots in a computationally efficient manner. The feature extraction algorithm was used to decompose the signal into a set of significant features which would be classified easily. Results obtained show that the algorithm was able to classify the damage state of the system with good accuracy. In the case of damage in both metal and composite specimens, the average classification in every case was above 84 per cent. The binary tree framework presented is computationally efficient because it reduces the number of comparisons needed and a smaller amount of data is required for training subsequent classifiers. A study on the computational efficiency of this scheme showed that the binary tree approach was 34 per cent more efficient than the 'one versus one' approach and 44 per cent more efficient than the 'one versus all' approach for the five class case. In this research, certain data sets used for classification had significant overlap because the induced damage in the structure caused only small changes in the sensor response. In the case of the lug joint a histogram was presented showing the extent of this overlap. This led to some amount of misclassification that was observed for certain classifiers. Since the performance of classifiers in lower nodes of the binary tree

is heavily dependent on the performance of classifiers higher up in the tree, the performance of this scheme is affected as a whole. The ROC curve presented for each classifier in the tree shows a good level of performance for each of the individual classifiers.

ACKNOWLEDGEMENTS

This research was supported in part by the MURI Program, Air Force Office of Scientific Research, grant number: FA9550-06-1-0309; Technical Monitor, Dr Victor Giurgiutiu and the NASA IVHM Program, grant number: NNX07AD70A, Program Manager, Dr Steve Arnold.

REFERENCES

1 **Yu, Y.-H., Choi, J.-H., Kweon, J.-H.,** and **Kim, D.-H.** A study on the failure detection of composite materials using an acoustic emission. *Compos. Structs*, 2006, **75**, 163–169.

2 **Genest, M., Martinez, M., Mrad, N., Renaud, G.,** and **Fahr, A.** Pulsed thermography for non-destructive evaluation and damage growth monitoring of bonded repairs. *Compos. Structs*, 2009, **88**(1), 112–120.

3 **Şimşir, M.** and **Ankara, A.** Comparison of two non-destructive inspection techniques on the basis of sensitivity and reliability. *Mater. Des.*, 2007, **28**(5), 1433–1439.

4 **Aymerich, F.** and **Meili, S.** Ultrasonic evaluation of matrix damage in impacted composite laminates. *Compos. Part B: Eng.*, 2000, **31**(1), 1–6.

5 **Kudela, P., Ostachowicz, W.,** and **Zak, A.** Damage detection in composite plates with embedded PZT transducers. *Mech. Syst. Signal Process.*, 2008, **22**, 1327–1335.

6 **Majumder, M., Gangopadhyay, T. K., Chakraborty, A. K., Dasgupta, K.,** and **Bhattacharya, D. K.** Fibre Bragg gratings in structural health monitoring – present status and applications. *Sens. Actuators A: Phys.*, 2008, **147**(1), 150–164.

7 **Coelho, C. K., Das, S., Chattopadhyay, A., Papandreou-Suppappola, A.,** and **Peralta, P.** Detection of fatigue cracks and torque loss in bolted joints. *Health Monit. Struct. Biol. Syst.*, 2007, pp. 653204–653212 (SPIE, San Diego, California, USA, 2007).

8 **Vapnik, V. N.** *Statistical learning theory*, 1998 (Wiley-Interscience, New York).

9 **Gao, G.-H., Zhang, Y.-Z., Zhu, Y.,** and **Duan, G.-H.** Hybrid support vector machines-based multi-fault classification. *J. China Univ. Min. Technol.*, 2007, **17**(2), 246–250.

10 **Shon, T.** and **Moon, J.** A hybrid machine learning approach to network anomaly detection. *Inf. Sci.*, 2007, **177**(18), 3799–3821.

11 **Hao, P.-Y., Chiang, J.-H.,** and **Tu, Y.-K.** Hierarchically SVM classification based on support vector clustering method and its application to document categorization. *Exp. Syst. Appl.*, 2007, **33**, 627–635.

12 **Yuan, S.-F.** and **Chu, F.-L.** Support vector machines-based fault diagnosis for turbo-pump rotor. *Mech. Syst. Signal Process.*, 2006, **20**, 939–952.

13 **Mallat, S. G.** and **Zhang, Z.** Matching pursuits with time–frequency dictionaries. *IEEE Trans. Signal Process.*, 1993, **41**(12), 3397–3415.

14 **Schölkopf, B.** and **Smola, A. J.** *Learning with kernels: support vector machines, regularization, optimization, and beyond*, 2002 (MIT Press, Cambridge, MA).

15 **Chattopadhyay, A., Das, S.,** and **Coelho, C. K.** Damage diagnosis using a kernel-based method. *Insight – Non-Destr. Test. Cond. Monit.*, 2007, **49**(8), 451–458.

16 **Coelho, C. K., Das, S., Chattopadhyay, A., Papandreou-Suppappola, A.,** and **Peralta, P.** Detection of fatigue cracks and torque loss in bolted joints (Ed. T. Kundu). In Proceedings of the SPIE, Smart Structures and Materials & Nondestructive Evaluation and Health Monitoring, SPIE, 2007, pp. 653204 (653201–653212)

17 **Cristianini, N.** and **Shawe-Taylor, J.** *An introduction to support vector machines: and other kernel-based learning methods*, 2000 (Cambridge University Press, Cambridge, UK).

18 **Vera-Candeas, P., Ruiz-Reyes, N., Rosa-Zurera, M., Lopez-Ferreras, F.,** and **Curpian-Alonso, J.** New matching pursuit based sinusoidal modelling method for audio coding. *IEE Proc. Vision, Image Signal Process.*, 2004, **151**(1), 21–28.

19 **Kovvali, N., Das, S., Chakraborty, D., Cochran, D., Papandreou-Suppappola, A.,** and **Chattopadhyay, A.** Time-frequency-based classification of structural damage. In Proceedings of the 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Honolulu, Hawaii, 23–26 April 2007, AIAA-2007-2055.

20 **Papandreou-Suppappola, A.** and **Suppappola, S. B.** Analysis and classification of time-varying signals with multiple time–frequency structures. *IEEE Signal Process. Lett.*, 2002, **9**(3), 92–95.

21 **Ghofrani, S., McLernon, D. C.,** and **Ayatollahi, A.** Comparing gaussian and chirplet dictionaries for time–frequency analysis using matching pursuit decomposition. In Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology, ISSPIT 2003, Darmstadt, Germany, 2003, pp. 713–716.

22 **Chakraborty, D., Zhou, W., Simon, D., Kovvali, N., Papandreou-Suppappola, A., Cochran, D.,** and **Chattopadhyay, A.** Time–frequency methods for structural health monitoring Sensor. Signal and Information Processing (SenSIP) Workshop, Sedona, USA, May 2008.

23 **Zweig, M. H.** and **Campbell, G.** Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin. Chem.*, 1993, **39**(4), 561–577.