ORIGINAL PAPER

A data-based structural health monitoring approach for damage detection in steel bridges using experimental data

Bjørn T. Svendsen[1] · Gunnstein T. Frøseth[1] · Ole Øiseth[1] · Anders Rønnquist[1]

Abstract

There is a need for reliable structural health monitoring (SHM) systems that can detect local and global structural damage in existing steel bridges. In this paper, a data-based SHM approach for damage detection in steel bridges is presented. An extensive experimental study is performed to obtain data from a real bridge under different structural state conditions, where damage is introduced based on a comprehensive investigation of common types of steel bridge damage reported in the literature. An analysis approach that includes a setup with two sensor groups for capturing both the local and global responses of the bridge is considered. From this, an unsupervised machine learning algorithm is applied and compared with four supervised machine learning algorithms. An evaluation of the damage types that can best be detected is performed by utilizing the supervised machine learning algorithms. It is demonstrated that relevant structural damage in steel bridges can be found and that unsupervised machine learning can perform almost as well as supervised machine learning. As such, the results obtained from this study provide a major contribution towards establishing a methodology for damage detection that can be employed in SHM systems on existing steel bridges.

1 Introduction

For bridges, structural health monitoring (SHM) systems provide information regarding the state of the bridge condition, with the aim of increasing the economic and life-safety benefits through damage identification. Many highway and railway bridges in Europe and the US, which experience increasing demands with respect to traffic loads and intensity, are approaching or have exceeded their original design lives. A large part of these bridges are steel and composite steel–concrete bridges. Based on an overview and comprehensive investigation of the common damage types experienced by such bridges that are reported in the literature [1], it is found that most damages are caused by fatigue and most frequently occur in or below the bridge deck. With the large number of existing bridges in infrastructure, lifetime

extension is the preferred option for ensuring continuous operation. Consequently, there is a need for reliable SHM systems that can detect both local and global structural damages in such bridges.

SHM is referred to as the process of implementing an automated and online strategy for damage detection in a structure [2, 3]. There are two main approaches in SHM: model-based and data-based [4, 5]. In the model-based approach, a numerical finite element model is continuously updated based on new measurement data to identify damage. The data-based approach, however, builds a statistical model based on experimental data only and generally relies on, but are not limited to, machine learning algorithms for damage identification. Additionally, a hybrid approach to SHM can be made that takes principles from both the model-based and data-based approaches into consideration. The analysis of the distributions of damage-sensitive features by machine learning algorithms, either supervised or unsupervised learning algorithms, is referred to as statistical model development [6]. Here, and in the context of SHM, supervised learning refers to the situation where data are available from both the undamaged and damaged conditions of

✉ Bjørn T. Svendsen
 bjorn.t.svendsen@ntnu.no

1 Department of Structural Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway

the structure, whereas unsupervised learning refers to the situation where data are available only from the undamaged condition. For bridges in operation, data from both the undamaged and damaged conditions are rarely available, and consequently, unsupervised learning is often required. Furthermore, bridges are subjected to changes in operational and environmental conditions, which complicate the detection of structural damage. One of the fundamental challenges in SHM is the process of separating changes caused by operational and environmental conditions from changes caused by structural damage, referred to as data normalization [7]. However, prior to including data normalization, principal knowledge about the damage detection possibilities of existing steel bridges must be established by considering relevant structural damage during stable operational and environmental conditions.

Applications of statistical model development have received increasing attention in the technical literature in recent years. In the absence of data from actual bridges, numerical models or test structures are commonly applied [8–11]. A significant contribution is the work performed by Figueiredo et al. [12], where damage detection under varying operational and environmental conditions is taken into consideration using the Los Alamos National Laboratory (LANL) test structure. Further work using the same test structure is performed by Santos et al. [13]. There are, however, inherent uncertainties in using a numerical model for damage detection, not only in the establishment of the numerical model itself but also in the modelling of structural damage. Similarly, damage introduced to laboratory test structures must be based on several assumptions and can, at best, be only a moderate representation of actual structural damage. Nevertheless, in the presence of data from actual bridges, important contributions to statistical model development related to damage detection studies on the Z24 prestressed concrete bridge [14] have been reported [15–18]. These studies mainly investigate the effects of operational and environmental conditions on damage detection using natural frequencies as damage-sensitive features. However, the relevant structural damage introduced to this bridge is mostly applicable to concrete bridges. Similar works with other bridge applications that are considered important contributions within this topic are found in Refs. [19, 20]. Except for [14, 21–27], few experimental studies have been reported in the literature where relevant structural damage is imposed on bridges. Furthermore, there are currently no studies in the literature where statistical model development is performed based on experimental studies on steel bridges.

This paper presents a data-based SHM approach for damage detection in steel bridges. The aim is to detect relevant structural damage based on the statistical model development of experimental data obtained from different structural state conditions. The Hell Bridge Test Arena, a steel riveted truss bridge formerly in operation as a train bridge and therefore representative of the many bridges still in service, is used as a full-scale damage detection test structure [28, 29]. An extensive experimental study is performed to obtain acceleration time series from the densely instrumented bridge under different structural state conditions during stable operational and environmental conditions. As such, relevant structural damage is implemented, the bridge is excited using a modal vibration shaker for simulating ambient vibration, and measurements are obtained from a setup containing two sensor groups to capture both the local and global responses of the bridge. The damage types chosen, including their locations, are based on the most common and frequently reported damage types in the literature: fatigue damage occurring in and below the bridge deck. The damage is considered highly progressed, representing loose connections and large cracks that open and close under dynamic loading. Autoregressive (AR) parameters are used as damage-sensitive features. The use of AR parameters has proven to be beneficial mainly because they are sensitive to the nonlinear behaviour of damage [30], which is a typical behaviour resulting from fatigue damage. Statistical model development is performed by considering both supervised and unsupervised machine learning. Four supervised machine learning algorithms are applied to test the abilities of different machine learning algorithms to detect structural damage, learn the structure of the data and determine how well the different damage types can be classified. The Mahalanobis squared distance (MSD), shown to be a highly suitable data normalization approach in terms of strong classification performance and low computational effort [12], is implemented as an unsupervised machine learning algorithm by novelty detection. Finally, the performances of the machine learning algorithms are assessed via both receiver operating characteristics (ROC) curves and confusion matrices.

Two novel contributions to the field of SHM regarding its application to bridges are made in this paper. First, statistical model development provides insight into the performances of several supervised machine learning algorithms based on a unique dataset established from a real-world application and allows for a study on the detectability of different damage types. Although these results have limited practical significance since data from both undamaged and damaged conditions are rarely available for bridges in operation, such information is invaluable for the SHM process and in the design of SHM systems. Second, a comparison between supervised and unsupervised learning algorithms is made. The implication of this insight provides a major contribution towards establishing a methodology for damage detection that can be employed in SHM systems on existing steel bridges.

The outline of this paper is as follows. Section 2 provides a description of the experimental setup, including the

damage introduced to the bridge and the operational and environmental conditions experienced during the measurements. Section 3 describes the feature extraction and the process of selecting the appropriate AR model order in addition to giving a brief overview of the supervised and unsupervised machine learning algorithms applied in this study. Section 4 presents the utilized analysis approach and the results obtained from the statistical models developed using the supervised and unsupervised machine learning algorithms. Finally, Sects. 5 and 6 summarize the work, discuss the analysis results obtained and suggest further work.

2 Experimental study

2.1 Experimental setup

The Hell Bridge Test Arena, shown in Fig. 1, is used as a full-scale damage detection test structure. The structural system of the bridge is composed of two bridge walls, the bridge deck and the lateral bracing. Figure 2 shows a schematic overview of the bridge.

An instrumentation system from National Instruments consisting of three cRIO-9036 controllers was used to

Fig. 1 The Hell Bridge Test Arena
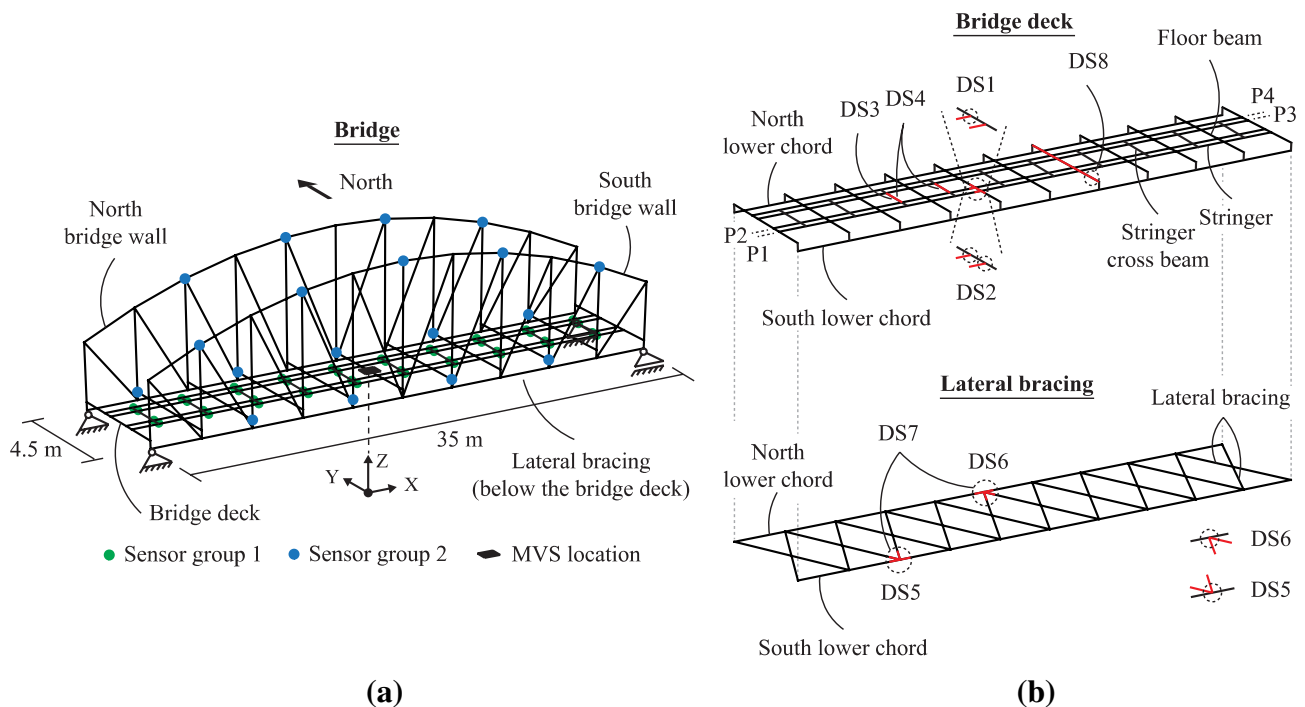
Fig. 2 Overview of the Hell Bridge Test Arena. **a** Bridge layout, including the locations of the modal vibration shaker (MVS) and sensors. **b** Damage introduced to the bridge deck and lateral bracing

acquire data from 58 accelerometers. The accelerometers were divided into a setup containing two sensor groups. Sensor group 1 consisted of 40 single-axis accelerometers (Dytran 3055D3) located below the bridge deck to measure the vertical response (global z-direction). Sensor group 2 consisted of 18 triaxial accelerometers (Dytran 3583BT and 3233A) located above the bridge deck, i.e. on the bridge walls, to measure the lateral and vertical responses (global y and z-directions). Data were sampled at 400 Hz. The data were detrended, filtered and resampled to 100 Hz before they were used for analysis. The bridge was excited in the vertical direction using a modal vibration shaker (APS 420) located at the bridge midspan. A band-limited random white noise in the range of 1–100 Hz with a maximum peak-to-peak excitation amplitude (stroke) of 150 mm was applied to simulate ambient vibration. The location of the modal vibration shaker and the sensor positions on the bridge are shown in Fig. 2a.

Ten different structural state conditions were considered, as summarized in Table 1. The structural state conditions were categorized into two groups: undamaged and damaged states. In the first group, the reference structural state of the bridge was represented by the baseline condition. Two baseline conditions were established under similar environmental conditions. In the second group, the damage states of the bridge were represented by different damage types with varying degrees of severity. Altogether, eight different damage states were established by considering four different damage types: stringer-to-floor beam connections; stringer cross beams; lateral bracing connections; and connections between floor beams and main load-carrying members. The variation in the degree of severity was considered by introducing each damage type at one or more locations in the bridge. The damage states were established consecutively in a sequence: damage was introduced, measurements were performed, and the damage was subsequently repaired. The damage types were chosen based on two considerations: first, these are the most common and frequently reported damage types in the literature [1]; and second, these are the most severe but relevant damage types for this type of bridge. An overview of the damage types introduced in the bridge, including their locations, is shown in Fig. 2b.

The damage types, shown in Fig. 3, were imposed by temporarily removing the bolts. All bolts were removed in each damage state condition. The damage types considered involved highly progressed damage, representing loose connections and large cracks that open and close under dynamic loading. As such, each damage represented a fully developed crack resulting from fatigue, leading to a total loss of functionality of the considered connection or beam. Such damage progression, which leads to a redistribution of forces, would

Table 1 Overview and descriptions of the structural state conditions

Label	State condition	Categorization[1]	Type	Description
UDS1	Undamaged	–	Baseline condition	Before all damage state conditions
UDS2	Undamaged	–	Baseline condition	After all damage state conditions
DS1	Damaged	Local	Stringer-to-floor beam connection	Single connection damaged
DS2	Damaged	Local	Stringer-to-floor beam connection	Multiple connections damaged
DS3	Damaged	Local	Stringer cross beam	Main part of single cross beam removed
DS4	Damaged	Local	Stringer cross beam	Main parts of multiple cross beams removed
DS5	Damaged	Global	Lateral bracing connection	Single connection damaged
DS6	Damaged	Global	Lateral bracing connection	Single connection damaged
DS7	Damaged	Global	Lateral bracing connection	Multiple connections damaged
DS8	Damaged	Global	Connection between floor beam and main load-carrying member	Single connection damaged

[1]Local: damage to the secondary steel structure. Global: damage to the primary steel structure

Fig. 3 The damage types imposed on the bridge. **a** Stringer-to-floor beam connection. **b** Stringer cross beam. **c** Lateral bracing connection. **d** Connection between floor beam and main load-carrying member

(a)　　　(b)　　　(c)　　　(d)
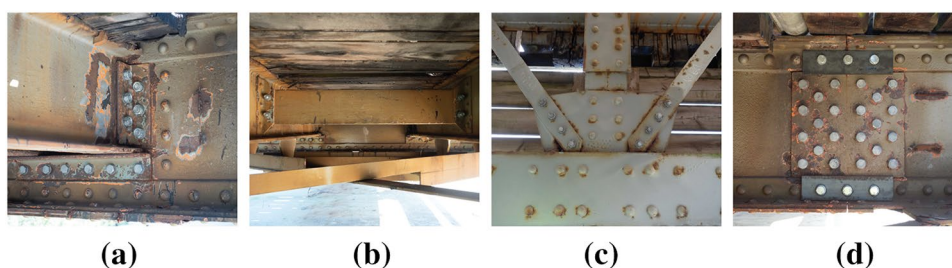
be demanding on the structure over time but not critical to the immediate structural integrity due to the redundancy inherent in the design of the bridge. Details regarding the damage types, including the associated mechanism, can be found in Ref. [1].

80 tests were performed for each state condition, resulting in a total of 800 tests. For each test, time series data were generated over 10.24 s, and data were obtained from a total of 75 sensor channels (40 channels from sensor group 1 and 35 channels from sensor group 2). It is noted that one channel in the lateral direction from sensor group 2 was excluded. Accelerometers were included in sensor groups 1 and 2 to capture the local and global responses of the bridge, respectively.

2.2 Operational and environmental conditions

Variability in the operational and environmental conditions imposes difficulties on the damage detection process. In general, operational conditions mainly include live loads, whereas environmental conditions include temperature effects, wind loading and humidity. During the measurements, no sources of variability were considered for the operational condition, which was limited to the operation of the modal vibration shaker only. The environmental conditions were logged during the measurements and are summarized in Table 2.

The environmental conditions were stable during the measurement period, which gives confidence that any changes observed in the results are caused by the damage imposed in the different damage states. An increase in the temperature occurred during the testing of DS3 and DS4; however, this temperature change was found to have little impact on the damage detection process.

Table 2 Environmental conditions during the measurement period

State	Temperature (°C)	Wind speed (m/s)	Weather
UDS1	10–11	1–2	Cloudy
UDS2	13	0–2	Partly cloudy and sunny
DS1	14	0–2	Partly cloudy and sunny
DS2	13	0	Partly cloudy and sunny
DS3	21	0–2	Sunny
DS4	22	0–2	Sunny
DS5	13	0–2	Partly cloudy and sunny
DS6	13	0–2	Partly cloudy and sunny
DS7	13	0–2	Partly cloudy and sunny
DS8	11	0–2	Partly cloudy and sunny

3 Feature extraction and machine learning algorithms

3.1 Feature extraction

In the context of SHM, machine learning is applied to associate the damage-sensitive features derived from measured data with a state of the structure; the basic problem is to distinguish between the undamaged and damaged states. In this study, an AR model is used to extract damage-sensitive features from time series data. For a specific time series, the AR(p) model of order p is given as [31]

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} + \varepsilon_t, \tag{1}$$

where y_t is the measured response signal, ϕ_j denotes the AR parameter(s) to be estimated and ε_t is the random error (residual) at the time index t. The use of AR parameters has proven useful for SHM applications regarding civil infrastructure mainly for three reasons [12, 30]: first, the parameters are sensitive to the nonlinear behaviour of damage, which is a typical behaviour resulting from fatigue damage; second, feature extraction depends only on the time series data obtained from the structural response; and third, the implementation is simple and straightforward.

To determine the appropriate order for a time series, two model selection criteria are commonly used: the Akaike information criterion (AIC) and the Bayesian model criterion (BIC). The AR model with the lowest AIC or BIC value gives the optimal order p.

3.2 AR model order selection

To find a common AR model order that can be applied to all time series, a model selection evaluation is performed. Analyses are performed for a selection of time series by considering the AIC and BIC values obtained by AR(p) models of increasing order p. The root mean square (RMS) values of the residuals are also considered as a heuristic approach, where the residuals are the differences between the model's one-step-ahead predictions and the real values of the time series. The results are based on the average results of all sensor channels, which are obtained by performing analyses of 80 tests in the undamaged state condition (40 tests from UDS1 and UDS2, respectively). Analyses of the AR parameters are performed using the *statsmodels* module in Python [32].

Figure 4 shows the results obtained by considering the normalized AIC and BIC values in addition to the RMS values of the residuals. The optimal order, p, is determined by the lowest values or by the convergence point of the values
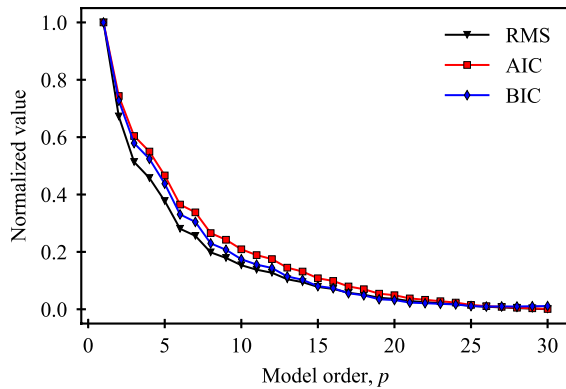
Fig. 4 Average AIC, BIC and RMS values for AR(p) models of increasing order p from all sensor channels for a selection of time series from the undamaged state condition

for a varying order. It is observed that the curves generally follow each other well and that the curves decrease rapidly for the lowest model orders. Although the AR models of order 8 or higher prove to be close to optimal representations of the time series, a lower-order model is used in this study. The choice of AR model order is important, not only because the optimal order provides the best representation of the time series but because the AR parameters are used as inputs for the machine learning algorithms in a concatenated format, affecting the dimension of the feature space. A high AR model order increases the dimension of the feature space. Consequently, there is a trade-off between obtaining the optimal time series representation and reducing the uncertainty of dealing with high-dimensional data. Data with high dimensionality may lead to problems related to the curse of dimensionality [33]; as the feature space dimension increases, the number of training samples required to generalize a machine learning model also increases drastically. Although a lower-order model is arguably not the optimal representation, it can still provide a good representation of the time series for use with all state conditions, and it reduces the uncertainty associated with high-dimensional data for the considered analysis approach. Therefore, for each test of each state condition, AR(5) models are established for the time series obtained from all sensor channels. The features are used as inputs to the machine learning algorithms for supervised and unsupervised learning.

3.3 Supervised learning

For supervised learning, a training matrix $\mathbf{X} \in \mathbb{R}^{n \times k}$ with n samples and k features and a test matrix $\mathbf{Z} \in \mathbb{R}^{m \times k}$ with m samples are composed of data from both the undamaged and damaged conditions. The target variable $\mathbf{y} \in \mathbb{N}^{n \times 1}$ is a vector composed of the class labels y. $\mathbf{x}_i \in \mathbb{R}^{1 \times k}$ and $\mathbf{z}_i \in \mathbb{R}^{1 \times k}$ denote arbitrary samples with an index i from the training and

test matrices, respectively, with a corresponding class label y_i. Note that k here is the total number of features to be included.

Supervised machine learning algorithms are considered for group classification only. The supervised machine learning algorithms are applied to test the abilities of different machine learning algorithms to detect structural damage, learn the underlying structure of the data and determine how well the different damage types can be classified. For this purpose, four common supervised machine learning algorithms are chosen: the k-nearest neighbours (kNN), the support vector machine (SVM), the random forests (RF) and the Gaussian naïve Bayes (NB) algorithms. Each algorithm yields different properties with respect to complexity, computational efficiency and performance in terms of the size of the training data and dimensionality of the feature space. In the context of SHM, the presented algorithms have never before been applied in supervised learning for damage detection on actual bridges, and only a few applications have been reported with respect to numerical studies or experimental studies using test structures [34, 35]. The theoretical backgrounds of the presented algorithms can be found in Refs. [36, 37], and practical examples (including implementations) can be found in Refs. [38, 39].

3.4 Unsupervised learning

For unsupervised learning, a training matrix $\mathbf{X} \in \mathbb{R}^{n \times k}$ with n samples and k features is composed of data from the undamaged condition only, and a test matrix $\mathbf{Z} \in \mathbb{R}^{m \times k}$ with m samples is composed of data from both the undamaged and damaged conditions. Unsupervised learning by novelty detection is implemented to take into consideration that data from the structure are generally only available in the undamaged condition. Novelty detection occurs when only training data from the undamaged state condition, i.e. the normal condition, are used to establish if a new sample point should be considered as different (an outlier). Consequently, if there are significant deviations, the algorithm indicates novelty. For SHM applications, several unsupervised machine learning algorithms have been reported in the literature [8, 12, 13, 17, 18, 34]. However, the MSD is found to be a highly suitable data normalization approach in terms of strong classification performance and low computational effort [12].

The MSD is a normalized measure of the distance between a sample point and the mean of the sample distribution and is defined as

$$DI_i = (\mathbf{z}_i - \overline{\mathbf{x}}) \mathbf{C}^{-1} (\mathbf{z}_i - \overline{\mathbf{x}})^T, \tag{2}$$

where $\mathbf{z}_i \in \mathbb{R}^{1 \times k}$ is the new sample point and potential outlier, $\overline{\mathbf{x}} \in \mathbb{R}^{1 \times k}$ is the mean of the sample observations (sample centroid) and $\mathbf{C} \in \mathbb{R}^{k \times k}$ is the covariance matrix. Both $\overline{\mathbf{x}}$ and $\mathbf{C}$ are obtained from the training matrix $\mathbf{X}$. The MSD is used as a damage index, which is denoted as DI.

To determine if a sample point is an inlier or outlier, a threshold value can be established using Monte Carlo simulations. The procedure of this method is adopted from Ref. [40] and summarized by the following steps for a 1% threshold:

1. A $(n \times k)$ matrix is constructed, where each element is a randomly generated number from a normal distribution with zero mean and unit standard deviation. The MSD for all n samples is calculated and the largest value is stored.
2. Step 1 is repeated for a minimum of 1000 trials. The array with all of the largest MSD values is structured in descending order.
3. The threshold value is established by considering the MSD values in the structured array in which 1% of the trials occur.

The threshold value is dependent on the numbers of observations and dimensions for the considered case.

4 Experimental analysis and results

4.1 Analysis approach

In the analysis approach used in this study, all sensor channels from each sensor group are taken into consideration simultaneously. The features are concatenated and applied as inputs for the machine learning algorithms. Hence, the setups for sensor groups 1 and 2 yield feature vectors with dimensions of 200 (40×5) and 175 (35×5), respectively. Statistical model development is performed based on the experimental study using both supervised and unsupervised learning for each setup separately.

4.2 Supervised learning

In supervised learning, labelled data are available. The supervised machine learning algorithms presented in Sect. 3.3 are implemented. The features are scaled to zero mean and unit variance for the algorithms, where relevant, to improve their performances.

The data are divided into 75% for training and 25% for testing. Consequently, the training matrix $\mathbf{X}$ has dimensions of 600×200 and 600×175 for sensor groups 1 and 2, respectively. The two baseline conditions (UDS1 and UDS2) are merged for the undamaged state condition (UDS) and include a total of 120 tests, whereas 60 tests are included for each damage state condition (DS1–DS8). Hence, a total of nine different class labels are included for evaluation. To find the optimal hyperparameters and increase the abilities of the machine learning algorithms to generalize to unseen data, a grid search with fivefold cross-validation is performed on the training data. The grid search conducts an exhaustive search over the specified hyperparameters to obtain the best cross-validation score. The hyperparameters of each algorithm, together with the final values obtained from the grid search, are summarized in Table 3. The test matrix $\mathbf{Z}$ has dimensions of 200×200 and 200×175 for sensor groups 1 and 2, respectively. Since the undamaged state condition is merged, it includes a total of 40 tests, whereas 20 tests are included for each damage state condition. For evaluation purposes, each sample from the test data is classified into one of the nine classes, resulting in a multi-class classification approach.

To evaluate the performances of the machine learning algorithms, also referred to as classifiers, ROC curves are established. The ROC curves represent the relative trade-offs between true positives (TP), or the probability of detection, and false positives (FP), or the probability of a false alarm

Table 3 Specified hyperparameters for the supervised machine learning algorithms

Algorithm	Hyperparameters				
	Parameter 1	Parameter 2	Parameter 3	Parameter 4	Parameter 5
SVM	Regularization strength (0.1, 0.01)	Kernel type (linear, linear)	Kernel coefficient (NA, NA)[1]	–	–
RF	Number of trees (200, 100)	Node impurity criterion (Entropy, Gini)	Maximum tree depth (10, 6)	Minimum leaf node samples (4, 4)	Maximum number of features for best split (14, 7)
NB[2]	–	–	–	–	–
kNN	Number of neighbours (1, 5)	Leaf size (1, 1)	Parameter for the Minkowski metrics (1, 1)	–	–

The final values of the hyperparameters are provided in parentheses for sensor groups 1 and 2. Details of the hyperparameters are specified in [38, 39]

[1] The kernel coefficient is not applicable (NA) for the linear kernel

[2] One parameter can be tuned in the NB algorithm; however, the default value was chosen

[41]. Figure 5 shows the averaged ROC curves, including the areas under the curves (AUCs), by considering all the classifiers for sensor groups 1 and 2. Each averaged ROC curve represents the micro-average of all classes, i.e. the contributions from all classes are aggregated to compute the average. To compare the classifiers, the ROC performances are reduced to scalar AUC values that represent the expected performances. Hence, a perfect classification is represented by the point (0, 1) with an AUC value of 1.0. Consequently, from the plots shown in Fig. 5, it is concluded that all classifiers perform well; however, the SVM outperforms the other classifiers and obtains perfect classification results for both sensor groups.

To obtain a more detailed view of the results, normalized multi-class confusion matrices for each classifier are established and shown in Fig. 6. By considering both sensor groups, it is clearly observed that the class labels are mostly incorrectly classified within each damage type, with a few exceptions. In particular, damage in the stringer cross beams, represented by the class labels DS3 and DS4, and damage in the lateral bracing connections, represented by DS5, DS6 and DS7, represent the majority of the incorrect predictions made. Accordingly, DS3 is incorrectly classified as DS4 or vice versa, and DS5 is incorrectly classified as DS6 or DS7, or vice versa. Interestingly, the stringer-to-floor beam connection type of damage, represented by DS1 and DS2, and the connection between floor beam and main load-carrying member type of damage, represented by DS8, are correctly classified by all classifiers. It is also observed that the naïve Bayes classifier performs worst for the stringer cross beam and lateral bracing connection types of damage relative to all classifiers, which is not very well reflected in the averaged ROC curves. There are two main explanations for the results obtained. First, damage in the stringer cross beams is categorized as local damage, and consequently, it is expected that introducing this type of damage generally has little influence on the structural response and thus makes it difficult to classify. Second, damage in the lateral bracing connections is categorized as global damage and is expected to have a global effect, rather than a local effect, on the structural response. Additionally, these two damage types influence the structural response mainly in the lateral direction of the bridge, which also explains the differences in the results obtained between sensor groups 1 and 2, since sensor group 1 only measures the response in the vertical direction.

An important part of the statistical model development process for supervised learning is to characterize the type of damage that can best be detected. Information regarding the damage types that are most difficult to detect when using the analysis approach considered is valuable information for bridge owners. In doing so, the true positive rate (TPR) (or recall) and positive predictive value (PPV) (or precision) are good measures. These measures are defined as

$$TPR = \frac{TP}{TP + FN}, \tag{3}$$

$$PPV = \frac{TP}{TP + FP}, \tag{4}$$

where TP denotes the true positives, FN represents the false negatives and FP signifies the false positives for a specified class. The TPR measures the proportions of correctly identified positives, whereas PPV measures the proportions of positive results. Additionally, the $F1$ score is the harmonic mean of the recall and precision. Table 4 summarizes the mean values of the TPR, PPV and $F1$ scores of the four classifiers for all class labels.
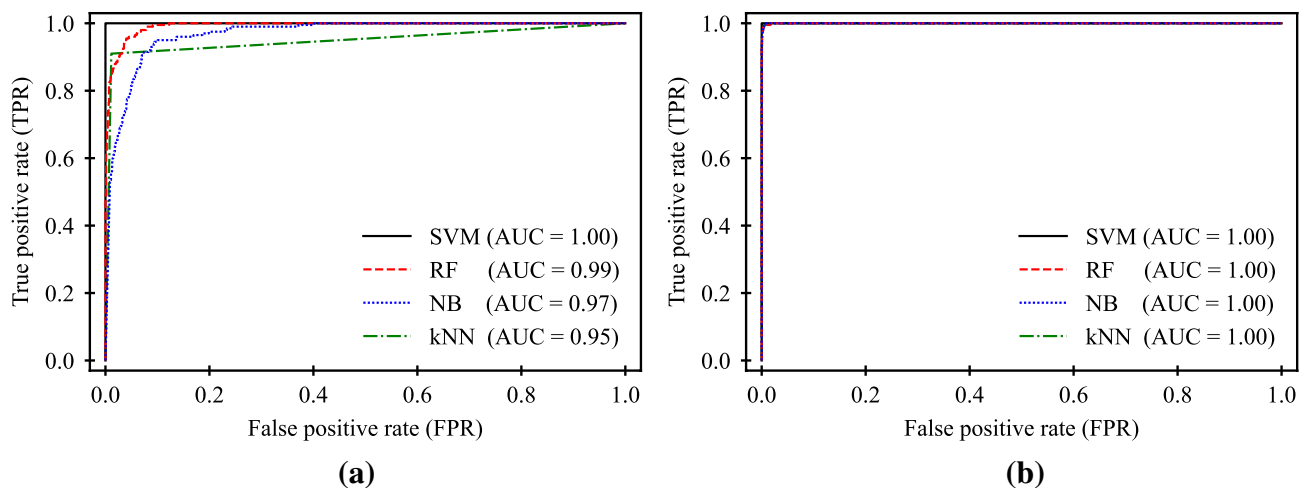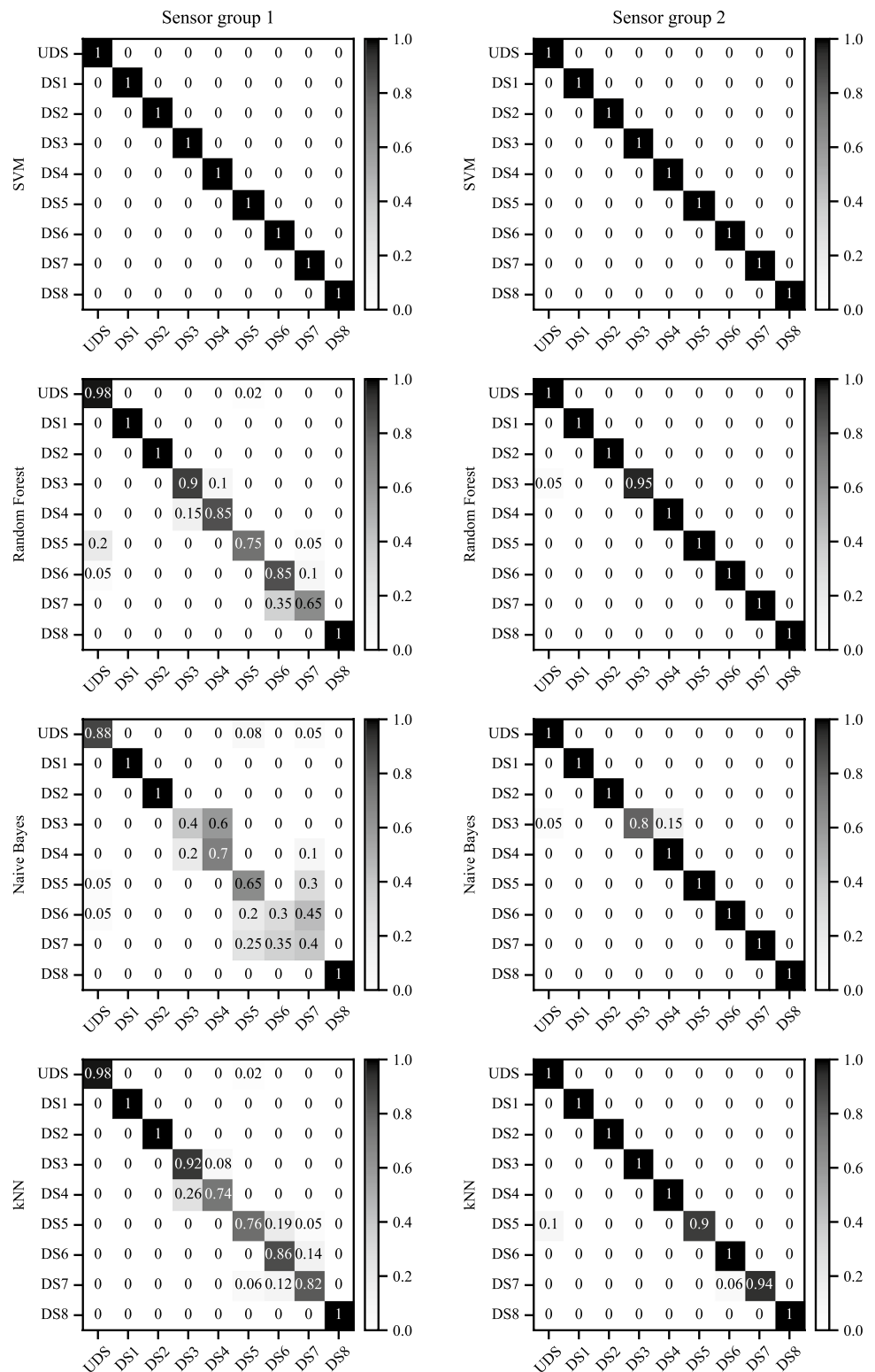

Fig. 5 Averaged ROC curves for the machine learning algorithms. **a** Sensor group 1. **b** Sensor group 2

Fig. 6 Normalized confusion matrices obtained from supervised learning by considering both sensor groups. For each confusion matrix, the predicted label is on the *x*-axis, and the true label is on the *y*-axis

From the results obtained, and by specifically considering the TPR, it is concluded that, on average, damage in the stringer-to-floor beam connections and the connection between floor beam and main load-carrying member are classified best, followed by damage in the stringer cross beams and the lateral bracing connections. Furthermore, a higher degree of stringer cross beam damage (DS4) is classified better than a lower degree (DS3); however, a similar conclusion cannot be made about the damage in the lateral bracing connections, with no obvious explanation.

Table 4 Mean values of the TPRs, PPVs and F1 scores of the four classifiers

Class label	Sensor group 1			Sensor group 2		
	TPR	PPV	F1 score	TPR	PPV	F1 score
UDS	0.957	0.958	0.956	1.000	0.988	0.994
DS1	1.000	1.000	1.000	1.000	1.000	1.000
DS2	1.000	1.000	1.000	1.000	1.000	1.000
DS3	0.805	0.836	0.811	0.938	1.000	0.966
DS4	0.822	0.827	0.820	1.000	0.967	0.983
DS5	0.790	0.837	0.808	0.976	1.000	0.988
DS6	0.752	0.709	0.722	1.000	0.983	0.991
DS7	0.718	0.733	0.722	0.985	1.000	0.992
DS8	1.000	1.000	1.000	1.000	1.000	1.000

Nevertheless, and most importantly, by using an appropriate classifier, such as the SVM, all damages can be correctly classified. The results from the SVM, based on the grid search, were obtained using a linear kernel for both sensor groups, proving that the dataset is linearly separable.

4.3 Unsupervised learning

In unsupervised learning, no labelled data are available. The MSD algorithm presented in Sect. 3.4 is implemented as a novelty detection method.

To obtain a reasonable amount of data from the undamaged state condition, the training matrix $\mathbf{X}$ is based on the measured data with a level of noise added. Thus, to establish the training matrix, 80 tests from the undamaged state condition are used as basic training data, i.e. 40 tests from each of the two baseline conditions (UDS1 and UDS2). The basic training data are copied 20 times, and each copy is subsequently corrupted with white Gaussian noise. Noise is added with a signal-to-noise ratio (SNR) equal to 20 using the definition

$$\text{SNR} = \frac{\mu}{\sigma}, \tag{5}$$

where μ is the mean of the absolute value of each individual feature considering all the samples in the basic training data and σ is the standard deviation (or RMS) of the noise. Notably, the statistics of each feature in the feature vector are taken into consideration during noise generation. The result is a suitable mean vector $\overline{\mathbf{x}}$ and covariance matrix $\mathbf{C}$ representing the undamaged state condition. Adding training data corrupted with noise improves and stabilizes the MSD algorithm and is found to be a good solution in situations where adequate measured training data are not available [40]. Consequently, the training matrix $\mathbf{X}$ has dimensions of 1600×200 and 1600×175 for sensor groups 1 and 2, respectively. The test matrix $\mathbf{Z}$ is composed of the remaining data from the undamaged state condition (UDS), i.e. 80 tests from the two baseline conditions (UDS1 and UDS2), and 80

tests from each of the damage state conditions (DS1–DS8). Thus, the test matrix has dimensions of 720×200 and 720×175 for sensor groups 1 and 2, respectively. For evaluation purposes, each sample from the test data is classified as either undamaged or damaged, resulting in a binary classification approach.

Figure 7 shows the ROC curves for the MSD algorithm, including the AUC values, obtained by considering both sensor groups. From this figure, it is concluded that the algorithm performs well: good classification results are obtained for sensor group 1, and perfect classification results are obtained for sensor group 2. The performance of the classifier regarding the 1% threshold is shown in Fig. 8, where the damage indices (DIs) from the state conditions are established for both sensor groups. Additionally, the state conditions of the respective test numbers are added for informative purposes at the top of the plots. In the binary classification approach, *FP* (false positive indications of damage) and *FN* (false negative indications of damage) are referred to as Type I and Type II errors, respectively. Type I errors are observed in all the tests based on the undamaged state condition (black markers) that are above the threshold,
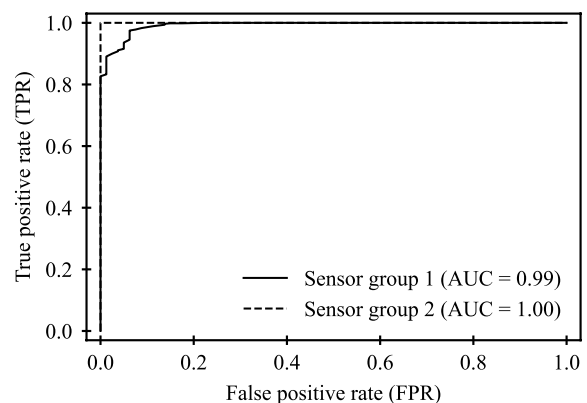
Fig. 7 ROC curves yielded by the MSD algorithm by considering both sensor groups
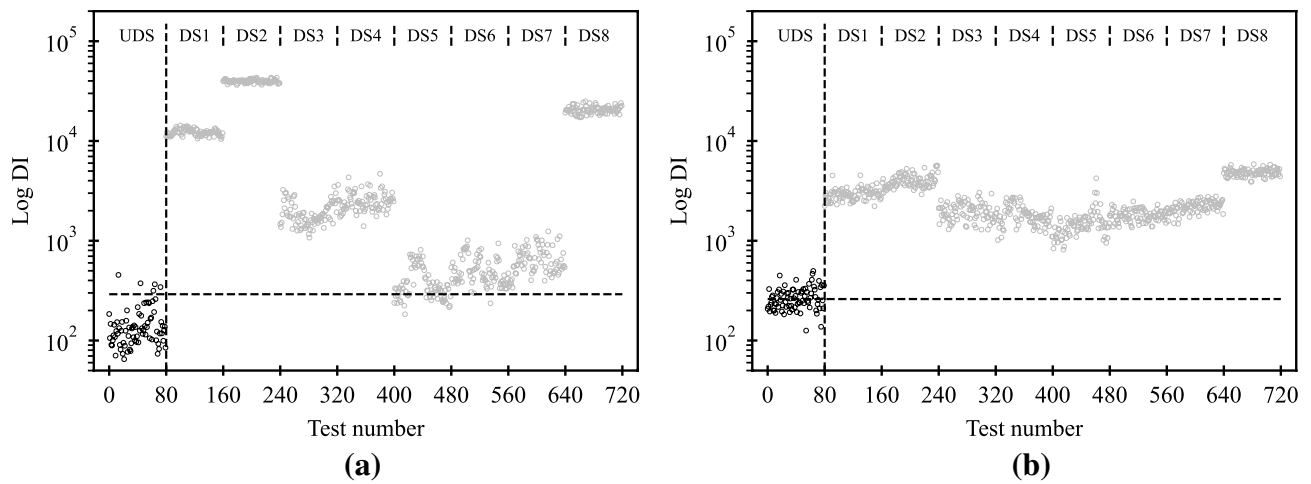
Fig. 8 Damage indices (DIs) yielded by the MSD algorithm for the undamaged (black) and damaged (grey) state conditions with respect to the threshold value (dashed line) established from the Monte Carlo simulations. **a** Sensor group 1. **b** Sensor group 2

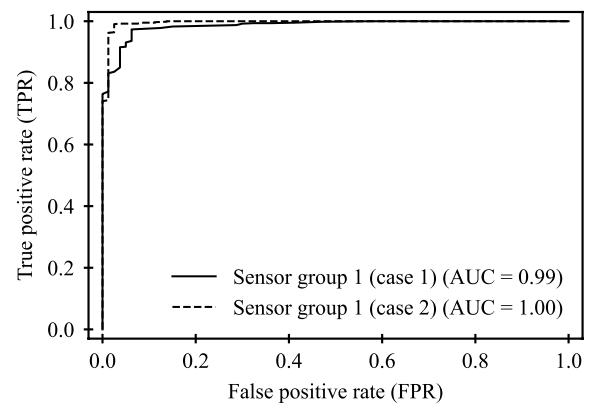**Table 5** Type I and II errors for both sensor groups yielded by the MSD algorithm

Setup	Error		
	Type I (FP)	Type II (FN)	Total
Sensor group 1	5 (6.3%)	26 (4.1%)	31 (4.3%)
Sensor group 2	39 (48.8%)	0 (0.0%)	39 (5.4%)

Fig. 9 ROC curves yielded by the MSD algorithm by considering both cases in sensor group 1

whereas Type II errors are observed in all the tests based on the damaged state condition (grey markers) that are below the threshold. For quantification purposes, these errors are summarized in Table 5 for both sensor groups with respect to the considered thresholds.

The results show that the setup for sensor group 1 yields the best performance with regard to avoiding false positive indications of damage (6.3%). The setup for sensor group 2 has a very low performance regarding false positive indications of damage (48.8%); however, it has an excellent performance when detecting damage (0.0%). Overall, the total misclassification rate obtained is lower for sensor group 1 (4.3%) than for sensor group 2 (5.4%). Nonetheless, these are good results for a novelty detection method. Additionally, an interesting observation in Fig. 8a is the clear trend showing that larger degrees of severity for each damage type also provide generally larger damage indices: DS2 > DS1, DS4 > DS3 and DS7 > DS6 and DS5. Although this is not observed in Fig. 8b, it clearly shows that the MSD algorithm performs well, particularly for the sensor group 1 setup.

4.4 Sensitivity analysis

To investigate the effect of a dense sensor network on the classification results, a sensitivity analysis is performed in the unsupervised learning with a reduced number of sensors in sensor group 1. Two cases are considered. In case 1, the sensors located in positions P1 and P2 in the longitudinal direction of the bridge are included, shown in Fig. 2, whereas the sensors in positions P3 and P4 are included in case 2. As such, only 20 sensors are included in each case, and the sensors located closest to the damage state conditions DS1, DS2 and DS8 are included in case 1. Figures 9 and 10 show the ROC curves and performance of the classifier regarding the 1% threshold for both cases, respectively. The Type I and II errors are summarized in Table 6.

From the results obtained, and through a comparison with the original case in sensor group 1 where all sensors are included, the effect of a dense sensor network can be clearly observed. Particularly, two important observations are made. First, although the performance regarding false negative indications of damage for case 1 (22.0%) is higher than that
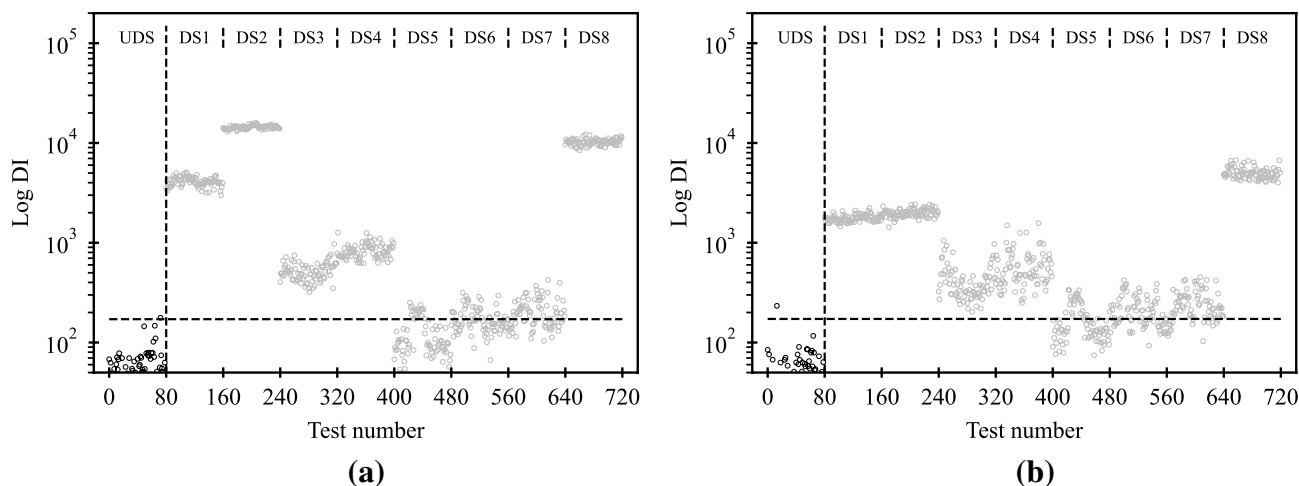
Fig. 10 Damage indices (DIs) yielded by the MSD algorithm for the undamaged (black) and damaged (grey) state conditions with respect to the threshold value (dashed line) established from the Monte Carlo simulations. **a** Case 1 in sensor group 1. **b** Case 2 in sensor group 1

Table 6 Type I and II errors for both cases in sensor group 1, including the original case, yielded by the MSD algorithm

Case	Error		
	Type I (FP)	Type II (FN)	Total
Original	5 (6.3%)	26 (4.1%)	31 (4.3%)
Case 1	1 (1.3%)	141 (22.0%)	142 (19.7%)
Case 2	1 (1.3%)	101 (15.8%)	102 (14.2%)

for case 2 (15.8%), case 1 shows more distinct results for damage state conditions DS1, DS2 and DS8 than case 2. These results indicate the positive effect of sensors being located near the damage locations. Second, the total misclassification rates of cases 1 (19.7%) and 2 (14.2%) are higher than that in the original case (4.3%). The low performance of both cases regarding outliers, or false negative indications of damage, is more unfavourable than that of incorrectly diagnosing inliers, or false positive indications of damage. Hence, reducing the number of sensors drastically increases the false negative indications of damage, which clearly demonstrates the importance of a dense sensor network.

5 Summary and discussion

From the statistical model development in supervised learning, three major observations are made. First, relevant structural damage in steel bridges can, in fact, be found. More specifically, by considering the hierarchical structure of damage identification [3, 42], level I (existence) and level III (type) damage detection can be performed. Second, how well structural damage can be classified strongly depends on the machine learning algorithm being applied. The machine

learning algorithms perform differently; however, the SVM achieves perfect classification results and outperforms the other algorithms, including those that can capture complex nonlinear behaviours. Third, a study on the detectability of the different damage types, based on the average performances of the classifiers, shows that damage to the lateral bracing connections is most often misclassified, followed by damage to the stringer cross beams. The other damage types are perfectly classified and represent the damage types that can best be detected. Such information is invaluable for the design of SHM systems with respect to instrumentation setup and sensor placement, and it provides insight into which damage types can generally be expected to be difficult to detect and those that are not for a similar analysis approach.

Data from both the undamaged and damaged conditions are generally not available for bridges in operation. Hence, unsupervised learning is required, where data from only the undamaged condition are available. Consequently, the results obtained from unsupervised learning are emphasized in this study. From the statistical model development, it is observed that the MSD algorithm performs well for the considered analysis approach with respect to minimizing the number of both false positive and false negative indications of damage. This algorithm has a high classification performance, is computationally efficient, and does not require any assumptions or tuning of parameters, although training data are added numerically to obtain improved performance. As such, the results obtained in this study, particularly by comparing supervised and unsupervised learning, are of major importance. Specifically, the unsupervised learning algorithm performs almost equally as well as the best supervised learning algorithms. This not only demonstrates that relevant structural damage in steel bridges can be classified using

unsupervised learning but also shows that a methodology for damage detection can be employed on existing bridges.

In this study, variability in the operational and environmental conditions was not taken into consideration using data normalization. To include data normalization, a larger variability in the environmental conditions than that experienced during the measurements is needed for the baseline data that represent the undamaged state condition. As such, periodic measurements that take all seasonal variations into consideration should be a minimum requirement. This is not included in the scope of this study, and future work should assess whether changes caused by damage can be separated from changes caused by any operational and environmental conditions. Such an assessment will reduce the uncertainty in the process of damage state assessment. Furthermore, it is acknowledged that for a future assessment based on long-term monitoring where data normalization is included, unsupervised learning methods, such as the auto-associative neural network (AANN) and Gaussian mixture models (GMMs), should be considered, particularly in the presence of nonlinear effects caused by variability in the operational and environmental conditions [17].

There are several advantages of the analysis approach presented in this study. The approach is computationally efficient from a machine learning perspective because many sensors can be evaluated simultaneously. Consequently, both supervised and unsupervised learning (novelty detection) can easily be performed. Furthermore, the approach allows for level I (existence) and level III (type) damage detection (the latter is only for supervised learning) but limits the possibility of performing level II (location) damage detection in unsupervised learning due to the large number of sensors widely distributed over the structure. The importance of a large number of sensors in a dense sensor network is demonstrated through a sensitivity analysis. The dense sensor network ensures a low total misclassification rate and enhances the damage detection capabilities. The main disadvantage of the analysis approach, however, is the high dimension of the feature space, which limits the number of features that can be included per sensor channel. From a general perspective, this limitation can be solved by increasing the number of samples in the training data or by performing dimension reduction techniques. From the results obtained in this study, despite this limitation, damage detection can still be successively performed.

AR parameters were found to be excellent damage-sensitive features in this study, and an AR(5) model was considered adequate. A low-order model was primarily used to reduce the dimension of the feature space for the analysis approach considered but also because low variations in the environmental and operational conditions were experienced during the measurements. The AR(5) model was able to capture the underlying dynamics of the structure and to accurately discriminate the undamaged and damaged states when applied as inputs for the machine learning algorithms under both supervised and unsupervised learning. However, it is important to note that environmental and operational conditions can introduce changes in the structural response. Furthermore, such changes can mask changes in any responses related to damage when a low model order is used, as shown by Figueiredo et al. [30]. This awareness is needed when using low-order models. Increasing the model order can easily be done when appropriate training data are available.

As a final note, the dataset used in this study was a result of measurements obtained using a dense instrumentation setup that was widely distributed over the structure, to capture both the local and global structural responses. Additionally, different damage types and degrees of severity were considered, and the operational and environmental conditions were logged. Consequently, the dataset obtained is unique in the context of performing damage detection in steel bridges.

6 Conclusion

The primary motive of SHM is to design a system that minimizes false positive indications of damage for economic and reliability concerns and false negative indications of damage for life-safety issues. For bridges, such a system should primarily be considered in an unsupervised learning approach, where data from only the undamaged condition are available.

This paper presented a data-based SHM approach for damage detection in steel bridges. The results obtained from an extensive experimental study proved that relevant structural damage in steel bridges, which is typically caused by fatigue, can be established using unsupervised learning. As such, this study provides a major contribution towards establishing a methodology for damage detection that can be employed in SHM systems on existing steel bridges. Future work will assess data normalization, where changes caused by damage can be separated from changes caused by any operational and environmental conditions, to reduce the uncertainty in the resulting damage state assessment.

References

1. Haghani R, Al-Emrani M, Heshmati M (2012) Fatigue-prone details in steel bridges. Buildings 2(4):456–476. https://doi.org/10.3390/buildings2040456

2. Farrar CR, Worden K (2007) An introduction to structural health monitoring. Philos Trans R Soc A Math Phys Eng Sci 365(1851):303–315. https://doi.org/10.1098/rsta.2006.1928

3. Worden K, Dulieu-Barton JM (2004) An overview of intelligent fault detection in systems and structures. Struct Heal Monit 3(1):85–98. https://doi.org/10.1177/1475921704041866

4. Farrar CR, Worden K (2012) Structural Health Monitoring: A Machine Learning Perspective. Wiley, Hoboken

5. Barthorpe RJ (2010) On Model- and Data-Based Approaches to Structural Health Monitoring. The University of Sheffield

6. Farrar CR, Doebling SW, Nix DA (2001) Vibration-based structural damage identification. Philos Trans R Soc A Math Phys Eng Sci 359(1778):131–149. https://doi.org/10.1098/rsta.2000.0717

7. Sohn H, Worden K, Farrar CR (2002) Statistical damage classification under changing environmental and operational conditions. J Intell Mater Syst Struct 13(9):561–574. https://doi.org/10.1106/104538902030904

8. Neves AC, González I, Leander J, Karoumi R (2017) Structural health monitoring of bridges: a model-free ANN-based approach to damage detection. J Civ Struct Heal Monit 7(5):689–702. https://doi.org/10.1007/s13349-017-0252-5

9. Neves AC, González I, Karoumi R, Leander J (2020) The influence of frequency content on the performance of artificial neural network–based damage detection systems tested on numerical and experimental bridge data. Struct Heal Monit. https://doi.org/10.1177/1475921720924320

10. Pan H, Azimi M, Yan F, Lin Z (2018) Time-frequency-based data-driven structural diagnosis and damage detection for cable-stayed bridges. J Bridg Eng 23(6):04018033. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001199

11. Malekzadeh M, Atia G, Catbas FN (2015) Performance-based structural health monitoring through an innovative hybrid data interpretation framework. J Civ Struct Heal Monit 5(3):287–305. https://doi.org/10.1007/s13349-015-0118-7

12. Figueiredo E, Park G, Farrar CR, Worden K, Figueiras J (2011) Machine learning algorithms for damage detection under operational and environmental variability. Struct Heal Monit 10(6):559–572. https://doi.org/10.1177/1475921710388971

13. Santos A, Figueiredo E, Silva MFM, Sales CS, Costa JCWA (2016) Machine learning algorithms for damage detection: Kernel-based approaches. J Sound Vib 363:584–599. https://doi.org/10.1016/j.jsv.2015.11.008

14. De Roeck G (2003) The state-of-the-art of damage detection by vibration monitoring: the SIMCES experience. J Struct Control 10(2):127–134. https://doi.org/10.1002/stc.20

15. Reynders E, Wursten G, De Roeck G (2014) Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification. Struct Heal Monit 13(1):82–93. https://doi.org/10.1177/1475921713502836

16. Figueiredo E, Moldovan I, Santos A, Campos P, Costa JCWA (2019) Finite element-based machine-learning approach to detect damage in bridges under operational and environmental variations. J Bridg Eng 24(7):04019061. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001432

17. Figueiredo E, Cross E (2013) Linear approaches to modeling nonlinearities in long-term monitoring of bridges. J Civ Struct Heal Monit 3(3):187–194. https://doi.org/10.1007/s13349-013-0038-3

18. Santos A, Figueiredo E, Silva M, Santos R, Sales C, Costa JCWA (2017) Genetic-based EM algorithm to improve the robustness of Gaussian mixture models for damage detection in bridges. Struct Control Heal Monit 24(3):1–9. https://doi.org/10.1002/stc.1886

19. Oh CK, Sohn H, Bae I-H (2009) Statistical novelty detection within the Yeongjong suspension bridge under environmental and operational variations. Smart Mater Struct 18(12):125022. https://doi.org/10.1088/0964-1726/18/12/125022

20. Magalhães F, Cunha A, Caetano E (2012) Vibration based structural health monitoring of an arch bridge: from automated OMA to damage detection. Mech Syst Signal Process 28:212–228. https://doi.org/10.1016/j.ymssp.2011.06.011

21. Farrar CR et al (1994) Dynamic characterization and damage detection in the I-40 bridge over the Rio Grande

22. Dilena M, Morassi A (2011) Dynamic testing of a damaged bridge. Mech Syst Signal Process 25(5):1485–1507. https://doi.org/10.1016/j.ymssp.2010.12.017

23. Döhler M, Hille F, Mevel L, Rücker W (2014) Structural health monitoring with statistical methods during progressive damage test of S101 Bridge. Eng Struct 69:183–193. https://doi.org/10.1016/j.engstruct.2014.03.010

24. Farahani RV, Penumadu D (2016) Damage identification of a full-scale five-girder bridge using time-series analysis of vibration data. Eng Struct 115:129–139. https://doi.org/10.1016/j.engstruct.2016.02.008

25. Kim CW, Chang KC, Kitauchi S, McGetrick PJ (2016) A field experiment on a steel Gerber-truss bridge for damage detection utilizing vehicle-induced vibrations. Struct Heal Monit 15(2):174–192. https://doi.org/10.1177/1475921715627506

26. Kim C-W, Zhang F-L, Chang K-C, McGetrick PJ, Goi Y (2021) Ambient and vehicle-induced vibration data of a steel truss bridge subject to artificial damage. J Bridg Eng 26(7):1–9. https://doi.org/10.1061/(asce)be.1943-5592.0001730

27. Maas S, Zürbes A, Waldmann D, Waltering D, Bungard V, De Roeck G (2012) Damage assessment of concrete structures through dynamic testing methods. Part 1 – Laboratory tests. Eng Struct 34:351–362. https://doi.org/10.1016/j.engstruct.2011.09.019

28. Svendsen BT, Frøseth GT, Rönnquist A (2020) Damage detection applied to a full-scale steel bridge using temporal moments. Shock Vib 2020:1–16. https://doi.org/10.1155/2020/3083752

29. Svendsen BT, Petersen ØW, Frøseth GT, Rønnquist A (2021) Improved finite element model updating of a full-scale steel bridge using sensitivity analysis. Struct Infrastruct Eng 10(1080/15732479):1944227

30. Figueiredo E, Figueiras J, Park G, Farrar CR, Worden K (2011) Influence of the autoregressive model order on damage detection. Comput Civ Infrastruct Eng 26(3):225–238. https://doi.org/10.1111/j.1467-8667.2010.00685.x

31. Box GEP, Jenkins GM, Reinsel GC, Ljung GM (2015) Time Series Analysis: Forecasting and Control (5th Edition). Wiley, Hoboken

32. Seabold S, Perktold J (2010) Statsmodels: Econometric and Statistical Modeling with Python. In: Proceedings of the 9th Python in Science Conference, pp 92–96. https://doi.org/10.25080/Majora-92bf1922-011.

33. Bellman RE (1961) Adaptive Control Processes. Princeton University Press, Princeton

34. Sun L, Shang Z, Xia Y, Bhowmick S, Nagarajaiah S (2020) Review of bridge structural health monitoring aided by big data and artificial intelligence: from condition assessment to damage detection. J Struct Eng 146(5):04020073. https://doi.org/10.1061/(asce)st.1943-541x.0002535

35. Flah M, Nunez I, Ben Chaabene W, Nehdi ML (2020) Machine learning algorithms in civil structural health monitoring: a

systematic review. Arch Comput Methods Eng. https://doi.org/10.1007/s11831-020-09471-9

36. Koutroumbas K, Theodoridis S (2008) Pattern Recognition, 4th edn. Academic Press Inc, Cambridge

37. Hastie T, Tibshirani R, Friedman J (2009) The Elements of Statistical Learning, 2nd edn. Springer, New York

38. Géron A (2019) Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. O'Reilly Media, Inc.

39. Pedregosa F et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830

40. Worden K, Manson G, Fieller NRJ (2000) Damage detection using outlier analysis. J Sound Vib 229(3):647–667. https://doi.org/10.1006/jsvi.1999.2514

41. Fawcett T (2004) ROC graphs: Notes and practical considerations for researchers. Palo Alto, CA

42. Rytter A (1993) Vibrational Based Inspection of Civil Engineering Structures. University of Aalborg, Denmark