Iranian Society of Acoustics and Vibration
The 12th ISAV2022 International Conference on Acoustics and Vibration
14-15 Dec 2022 Tehran - Iran
Amirkabir University of Technology

Vibration-based Damage Assessment in Environment Changes by Locally Unsupervised Learning

Mohammadreza Mahmoudkelayeh[a*], Behzad Saeedi Razavi [b]

[a] *PhD student, Faculty of Civil & Earth Resources Engineering,Central Tehran Branch, Islamic Azad University, Tehran, Iran.*

[b] *Research Assistant professor, Standard Research Institute, Technical and engineering faculty, Construction and mineral Department, Karaj, Iran.*

* *Corresponding author e-mail: mrmkelayeh@gmail.com*

Abstract

Health monitoring of civil structures by continuous measurements of vibration data is one of the effective and reliable techniques for ensuring structural safety and serviceability. Environment changes and the methodology for damage assessment are two important challenges that play critical roles in achieving outstanding results due to the emergence of serious errors. This research intends to propose a new vibration-based method under the idea of locally unsupervised learning for addressing the challenges. The proposed method consists of two steps of data clustering and damage assessment. In the first step, a new clustering algorithm called locally undirected-based graph density peak clustering (LUG-DPC) is presented to split dynamic features into pre-determined clusters and supply local information. The second step utilizes such local information to estimate local mean vectors and local covariance matrices that make the main elements of the anomaly detector based on the Mahalanobis distance. The major contributions of this paper contain developing an innovative machine learning-aided method and introducing the LUG-DPC for damage assessment Long-term continuous natural frequencies of a full-scale concrete bridge are used to verify the proposed method with some comparisons. Result indicates the proposed method can alleviate the environment effects and obtain reasonable results with inconsiderable errors compared to some well-known techniques.

Keywords: Structural health monitoring; environment variability; unsupervised learning; clustering.

1. Introduction

Damage is an adverse change in structures that can cause undesirable alterations in structural properties and vibration responses. Detection of damage is a necessity for structural systems for reasons of safety and economy. On this basis, structural health monitoring (SHM) undertakes this task by recording vibration measurements, either constructing a numerical model of the structure or extracting features from vibration measurements, and analysing the current state of the structure in terms of the occurrence of damage or normal operation [1].

Recently, machine learning has offered a new and practical technology for SHM in a data-based manner without any numerical model construction and updating. This technique is often categorized into supervised learning and unsupervised learning depending upon the type of data and the possibility of having supervision information (labelled data) [2, 3]. Because civil structures are often complex, large-scale, and expensive, it is illogical to impose intentional damage cases to provide fully labelled features and information and apply supervised learning techniques. For this reason, it is further preferable in civil engineering societies to take advantage of unsupervised learning [4-8].

In SHM applications via the concept of unsupervised learning, some challenging issues are still problematic. In a long-term continuous SHM program, the unlabelled features such as modal properties are influenced by different types of environment changes including daily and seasonal temperature fluctuations, humidity and moisture variations, and wind speed and direction. The demanding issue is that these variability conditions affect the mechanical and physical characteristics of a civil structure; that is, mass and stiffness. Under such circumstances, the structural behaviour and responses in these conditions may mistakenly interpreted as damage leading to a false alarm or false positive error. Alternatively, it is probable that the severity of environment changes may be larger than a real damage pattern, especially minor damage. In this case, it is accurately possible to alarm the occurrence of real damage and such changes mask the damage leading to a false detection or false negative error [9, 10].

Although many research studies were conducted to investigate any types of environment changes [11], predict their severities [12], and eliminate from vibration-based features, especially modal data [13], most of them focused on global statistical and unsupervised learning techniques. One of the drawbacks of global approaches is that those consider the whole features that are much more influenced by the variability conditions. In this case, it is possible to develop deceptive unsupervised learning models for SHM resulting in false positives and false negatives. To alleviate this drawback, one can take advantage of the idea of local unsupervised learning. This strategy is generally intended to provide local information from the whole data. Some local unsupervised learning techniques developed for SHM are based on local principal component analysis [14], unsupervised *k*-nearest neighbour [15], local partitioning [16, 17], unsupervised feature selection with local distance learning [18].

Based on the aforementioned challenges, we intend to propose a locally unsupervised learning method for SHM under environment changes. The central core of this method relates to two parts of local feature clustering and damage indicator determination. For the first part, this article proposes a new clustering algorithm called locally undirected graph-based density peak clustering (LUG-DPC) [19] for splitting the whole training features into pre-determined clusters. In the second part, the feature samples of each cluster are taken into account to estimate the local mean vector and local covariance matrix of that cluster for damage detection. On this basis, the Mahalanobis distance is applied to calculate distance values of training and test data points. Long-term natural frequencies of a full-scale concrete bridge are used to verify the correctness and trustworthy of the proposed method. Results indicate that the proposed method not only enables us to address the challenge of environment changes but also gives reliable SHM results.

2. Locally undirected graph-based density peak clustering

The LUG-DPC is a new clustering technique proposed Cheng et al. [19], who developed it to divide unlabelled data points in pre-determined clusters with arbitrary shapes. The central assumption on this technique is that data points can be clustered around centres or separated by regular geometric curves. The fundamental principle of this technique is based on an undirected graph that connects all the vertices with the minimum sum of edge weights and without any cycles or directions. From a machine learning aspect, a graph is a group of nodes and edges, where the nodes build the graph vertices that relate to original data points and the edges make connections between these points. Depending upon the type of connection, the graphs can be decomposed into undirected and directed categories.

With these descriptions, the LUG-DPC technique sets out to choose some data points as useful features in such a way that those can preserve the cluster shapes and ignore the useless points, which may be influenced by any source of variability such as environment changes. More precisely, this technique firstly attempts to find data points with the largest local density values, called here LDPs, among their neighbours and assigns the remaining points to the corresponding LDPs. Second, the LUG-DPC exploits a shared-neighbour-based distance to calculate the dissimilarity between the LDPs and their neighbours, in which case it is feasible to construct an undirected graph of the LDPs as the vertices and their neighbours as the edges of this graph. Third, it needs to segment the LDPs into pre-determined clusters. Finally, the assigned data points to their LDPs are accommodated into their clusters [19].

To find the LDPs and construct their undirected graph, it is essential to define the function of local density of data points. Given a data sample **x,** its local density is expressed as [19]:

$$\rho(\mathbf{x}) = \frac{n_\rho}{\sum_{\mathbf{y}\in\mathbf{N}_\mathbf{x}} d(\mathbf{x},\mathbf{y})} \tag{1}$$

where n_ρ denotes the number of nearest neighbours of **x**; $d(\mathbf{x},\mathbf{y})$ is the distance operation between the samples **x** and **y**, which the term "*d*" refers to the Euclidean distance function; $\mathbf{N}_\mathbf{x}$ represents a set containing the nearest neighbours of the sample **x**. It should be clarified that this set can be obtained from the general idea of natural neighbour proposed by Zhu et al. [20] aiming at finding the mutual neighbours between two samples.

The next step of the LUG-DPC is to find the LDPs under the concept of local density and the nearest neighbours of each data point. After determining the local density of each sample, the selection of the LDPs is a trivial task. For this goal, it only suffices to choose a point with the maximum local density among the original sample (**x**) and its nearest neighbours. In other words, a data point is a LDP if and only if its local density is maximum among all nearest neighbours. Subsequently, it is necessary to compute the distances between LDPs via the shared-neighbour-based distance for constructing the undirected graph. Let **x** and **y** denote two LDPs, the shared-neighbour-based distance between them is formulated as follows [19]:

$$\lambda = \begin{cases} \dfrac{d(\mathbf{x},\mathbf{y})}{n_d \sum_{\mathbf{v}\in\delta(\mathbf{x},\mathbf{y})} \rho(\mathbf{v})}, & if \ \ |\delta(\mathbf{x},\mathbf{y})| \neq 0 \\ \max(d)\left(1+d(\mathbf{x},\mathbf{y})\right), & if \ \ |\delta(\mathbf{x},\mathbf{y})| = 0 \end{cases} \tag{2}$$

where n_d is the number of the shared neighbours between two LDPs, which are the intersections of their neighbours and fall in $\delta(\mathbf{x},\mathbf{y})$; $\rho(\mathbf{v})$ denotes the local density of the sample **v**, which is one of the shared neighbours of both LDPs **x** and **y**. In Eq. (2), max(*d*) is the maximum distance between all pairs of the LDPs. Accordingly, the LDPs and shared-neighbour-based distance values are used to build the undirected graph. Based on the number of clusters, the LDPs are segmented into clusters of interest and receive cluster indices or labels. Eventually, the other (non-LDP) data points assign their cluster indices inspired by their LDPs. This means that the LUG-DPC technique

spreads all unlabelled (training) data points into pre-determined clusters and gives each of them a label, which is the number of its cluster.

3. Proposed SHM method

The proposed SHM method consists of two levels of feature clustering by the LUG-DPC technique, as explained in the previous section, and damage assessment by an indicator or anomaly detector based on the Mahalanobis-squared distance (MSD).

3.1 Feature clustering

In this level, the proposed method starts by determining the optimal cluster number (k). In this research, an approach via the Gap statistic [21] is considered. Suppose that $\mathbf{X}=[\mathbf{x}_1,\ldots,\mathbf{x}_n]$ is the training data (matrix) regarding the undamaged condition of the structure containing n data points (feature vectors) of t variables; that is, $\mathbf{x}_i=[x_1,\ldots,x_t]^{\mathrm{T}}$, where $i=1,\ldots,n$. Using a sample cluster, the Gap-based cluster determination algorithm examines different clusters and find one of them that can satisfy the Gap statistic as presented in Tibshirani et al. [21]. With this cluster number (k), in the s the LUG-DPC splits the n training data points into k clusters $\{\mathbf{C}_1,\ldots,\mathbf{C}_k\}$. It should be described that that each cluster is a matrix including some data points (i.e., $N_1,\ldots,N_k$) with t variables.

3.2 Damage assessment

The second level of the proposed method is based on developing an anomaly detector or damage indicator through the MSD metric. For this purpose, the clustered sets are incorporated to estimate their local mean vectors $\{\mathbf{m}_1,\ldots,\mathbf{m}_k\}$, each of which contains t variables, and local covariance matrices $\{\mathbf{S}_1,\ldots,\mathbf{S}_k\}$, each of which is a ($t\times t$) matrix. Note that the mean vectors and covariance matrices are the main elements of the MSD-based anomaly detector. Given the ith training feature $\mathbf{x}_i$, where $i=1,\ldots,n$, its damage indicator is written as follows:

$$DI\left(\mathbf{x}_i\right)=\min\left\{\left(\mathbf{x}_i-\mathbf{m}_k\right)^T\mathbf{S}_k^{-1}\left(\mathbf{x}_i-\mathbf{m}_k\right)\right\} \tag{3}$$

where $DI(\mathbf{x}_i)$ is the damage indicator of $\mathbf{x}_i$, which is the minimum distance quantise among all k distances. Similarly, one can obtain the damage indicator of each test point ($\mathbf{z}_j$), where $j=1,\ldots,m$ and m denotes the number of test points, in the following form:

$$DI\left(\mathbf{z}_j\right)=\min\left\{\left(\mathbf{z}_j-\mathbf{m}_k\right)^T\mathbf{S}_k^{-1}\left(\mathbf{z}_j-\mathbf{m}_k\right)\right\} \tag{4}$$

Having considered all damage indicators of the training samples $\{DI(\mathbf{x}_1),\ldots,DI(\mathbf{x}_n)\}$, the standard confidence interval is used to determine a threshold for damage detection. In this regard, it is expected that the damage indicators of the training data related to the undamaged state of the structure fall below the threshold. If the damage indicator of the test sample does not exceed the threshold, the proposed method declares that the structure is still undamaged; otherwise, it alarms the occurrence of damage.

4. Results and discussions

To validate the effectiveness and reliable performance of the proposed method, the long-term continuous natural frequencies of a post-tensioned concrete box-girder bridge called Z24 [22] are considered here. Figure 1(a) shows an actual picture of the Z24 bridge. This structure contained the main span of the length of 30m and two sides of 14m, as shown in Figure 1(b). In order to build a new structure with larger side span, the Z24 bridge was demolished at the end of 1998. Before this procedure, continuous monitoring tests were performed to quantify environment data (temperature) and record vibration responses (acceleration time histories). Progressive damage patterns were also defined to simulate realistic damaged states of the structure. Utilizing the recorded acceleration re-

sponses, an automated technique for operational modal analysis was considered to extract the modal properties, particularly long-term natural frequencies of four modes.

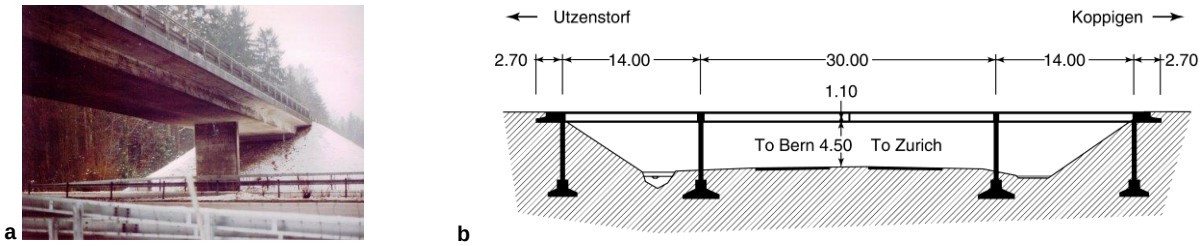

Figure 1. The Z24 bridge: (a) an actual picture, (b) the major dimensions of the side view

The dataset of the natural frequencies contained some missing values; hence, those are discarded to obtain a set of 3932 measurements of the natural frequencies as shown in Figure 2(a), where the first 3475 samples relate to the undamaged state of the bridge and the remaining points are concerned with the damaged state. As can be seen, the environment changes in the natural frequencies of the undamaged state are dominant and should be removed by the proposed method. For this objective, the training data should be made that is a matrix of 3128 natural frequencies of the undamaged state (n=3128), which is equivalent to 90% of all feature points related to this state. As such, the remaining natural frequencies of the undamaged condition and all features of the damaged state are comprised the test data leading to 804 test points (m=804). Note that the training and test samples (vectors) consist of four variables; that is, t=4.

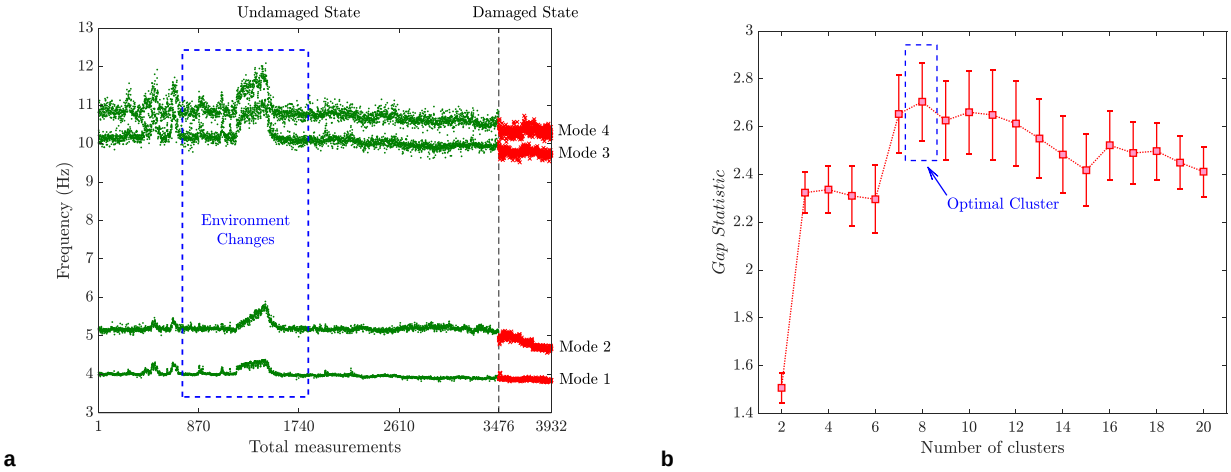

Figure 2. (a) The natural frequencies of the bridge, (b) optimal cluster selection by the Gap statistic

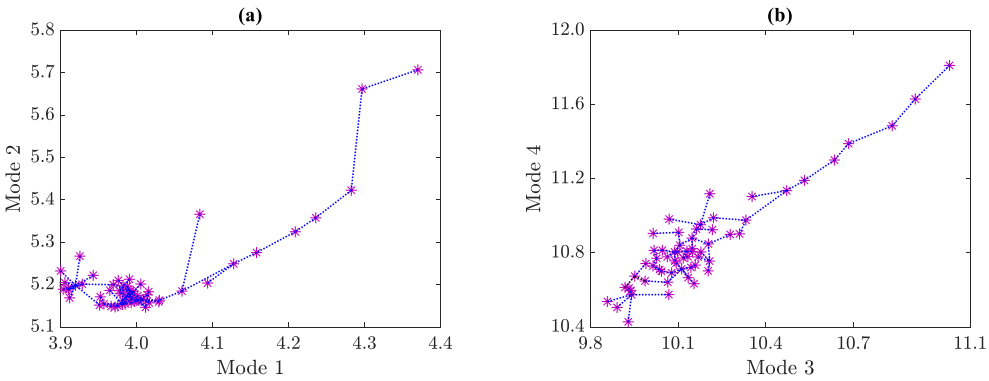

Figure 3. The LDPs of the training samples: (a) modes 1 and 2, (b) modes 3 and 4

The proposed method starts by finding the optimal cluster number by the Gap statistic. For this purpose, 19 sample clusters are considered from 2 to 20 to implement the clustering process. Figure 2(b) shows the Gap statistic values of these sample clusters, where the optimal number of clusters is equal to k=8. Hence, the 3128 training samples should be split into eight clusters. Furthermore, Figure 3 illustrates the undirected graphs of the selected LDPs concerning the training data points based on two successive modes. Totally, the LUG-DPC technique selects 60 samples of the training data as the LDPs; hence, the other 3068 points treat as the non-LDPs. According to the eight clusters, the LDPs receives their labels from 1-8. Subsequently, the non-LDP points join their LDPs and complete their clusters. Figure 4(a) displays the distribution of all training points (either LDPs and non-LDPs) into the eight clusters and also their label numbers.

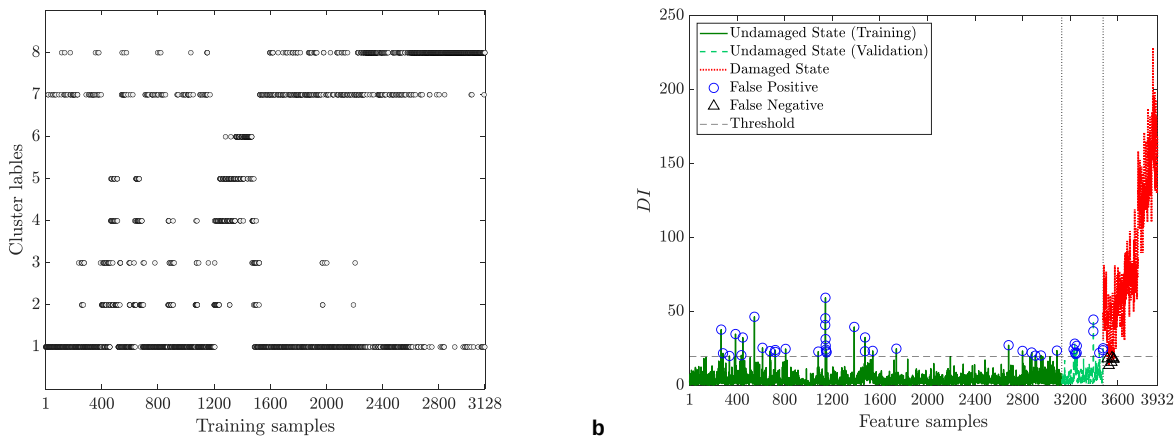

Figure 4. (a) Cluster labels of the training points, (b) damage assessment by the proposed method

Using the clusters $\mathbf{C}_1,\ldots,\mathbf{C}_8$, their mean vectors and covariance matrices are estimated to compute MSD values for damage assessment. The result of this procedure is shown in Figure 4(b), where the horizontal line refers to the threshold based on the standard confidence interval under 0.001 significance level. As can be observed, the first main important note in Figure 4(b) is that the proposed method with the help of the LUG-DPC could mitigate the major environment changes in the natural frequencies of the undamaged state. Moreover, it is seen that almost all distance values or damage indicators related to this state are below the threshold with the expectation of some points referring to false positives. The other important note in Figure 4(b) the majority of the damage indicators of the damaged state exceed the threshold except for a few points indicating inconsiderable false negatives. If the threshold is ignored, one can suggest that the proposed method could provide reliable detection accuracy due to easily discriminating the damaged state from the undamaged one.

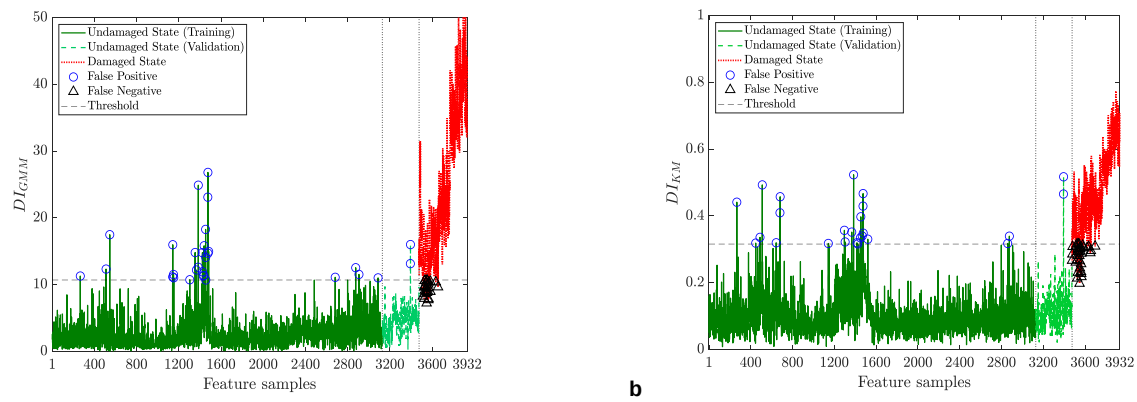

Figure 5. Comparative studies on damage assessment in the Z24 bridge: (a) GMM-MSD, (b) KMC-ESD

In order to better demonstrate the superiority of the proposed method, it is compared with some well-known unsupervised learning methods. Similarly, these approaches are based on feature clustering via the Gaussian mixture model (GMM) and *k*-means clustering (KMC) [12]. For the second part, the MSD and Euclidean-squared distance (ESD) are used to compute damage indicators based on the outputs of the GMM and KMC. The same Gap statistic and sample clusters are used to determine the optimal component and cluster numbers for the GMM and KMC, which correspond to 5 and 13, respectively. Figure 5 indicates the results of damage assessment via the GMM-MSD and KMC-ESD. It is obvious that the environment changes in the natural frequencies of the undamaged state are still available in the damage indicators of these techniques. Due to this event, large threshold values have been determined that lead to considerable false negatives. Moreover, there are many false positive errors in Figure 5. Without the thresholds, it is difficult to appropriately distinguish the damaged state from the undamaged one because some damage indicators of the undamaged condition are either equal to or larger than the corresponding indicators of the damaged state. Therefore, it can be concluded that the proposed method is superior to the well-known GMM-MSD and KMC-ESD techniques.

5. Conclusions

This paper proposed a new unsupervised learning method for vibration-based damage assessment in civil structures under environment changes. The proposed method contained feature clustering by the new technique called LUG-DPC and anomaly detection by the MSD. The first level undertook the segmentation of the whole training data into clusters to provide local information. In the second level, such information was considered to determine the main elements of the MSD-based anomaly detector; that is, the local mean vectors and local covariance matrices. Continuous natural frequencies of the Z24 bridge were applied to verify the proposed method. The results of this research indicated that the proposed local unsupervised learning with the aid of the LUG-DPC can address the major challenge of environment changes and obtain reliable results of damage assessment with high accuracy. The comparative studies showed that this method is superior to the well-known GMM-MSD and KMC-ESD techniques in terms of smaller error rates and higher detection accuracy.

REFERENCES

1. H.-N. Li, L. Ren, Z.-G. Jia, T.-H. Yi, and D.-S. Li, "State-of-the-art in structural health monitoring of large and complex civil infrastructures". *Journal of Civil Structural Health Monitoring* **6**(1), 3-16 (2016).
2. A. Malekloo, E. Ozer, M. AlHamaydeh, and M. Girolami, "Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights". *Structural Health Monitoring* **21**(4), 1906-1955 (2022).
3. H. Sarmadi and A. Entezami, "Application of supervised learning to validation of damage detection". *Archive of Applied Mechanics* **91**(1), 393-410 (2021).
4. H. Sarmadi and K.-V. Yuen, "Early damage detection by an innovative unsupervised learning method based on kernel null space and peak-over-threshold". *Computer-Aided Civil and Infrastructure Engineering* **36**(9), 1150-1167 (2021).
5. M. H. Daneshvar and H. Sarmadi, "Unsupervised learning-based damage assessment of full-scale civil structures under long-term and short-term monitoring". *Engineering Structures* **256**, 114059 (2022).
6. A. Entezami, H. Shariatmadar, and C. De Michele, "Non-parametric empirical machine learning for short-term and long-term structural health monitoring". *Structural Health Monitoring*, (2022).
7. H. Sarmadi, A. Entezami, M. Salar, and C. De Michele, "Bridge health monitoring in environmental variability by new clustering and threshold estimation methods". *Journal of Civil Structural Health Monitoring* **11**(3), 629-644 (2021).
8. A. Entezami, H. Shariatmadar, and S. Mariani, "Early damage assessment in large-scale structures by innovative statistical pattern recognition methods based on time series modeling and novelty detection". *Advances in Engineering Software* **150**, 102923 (2020).

9. H. Sarmadi and K.-V. Yuen, "Structural health monitoring by a novel probabilistic machine learning method based on extreme value theory and mixture quantile modeling". *Mechanical Systems and Signal Processing* **173**, 109049 (2022).
10. A. Entezami, S. Mariani, and H. Shariatmadar, "Damage detection in largely unobserved structures under varying environmental conditions an autoregressive spectrum and multi-level machine learning methodology". *Sensors* **22**(4), 1400 (2022).
11. Y. Zhou and L. Sun, "Effects of environmental and operational actions on the modal frequency variations of a sea-crossing bridge: A periodicity perspective". *Mechanical Systems and Signal Processing* **131**, 505-523 (2019).
12. H. Sarmadi, "Investigation of machine learning methods for structural safety assessment under variability in data: Comparative studies and new approaches". *Journal of Performance of Constructed Facilities* **35**(6), 04021090 (2021).
13. W.-H. Hu, Á. Cunha, E. Caetano, R. G. Rohrmann, S. Said, and J. Teng, "Comparison of different statistical approaches for removing environmental/operational effects for massive data continuously collected from footbridges". *Structural Control and Health Monitoring* **24**(8), e1955 (2017).
14. A.-M. Yan, G. Kerschen, P. De Boe, and J.-C. Golinval, "Structural damage diagnosis under varying environmental conditions—part II: local PCA for non-linear cases". *Mechanical Systems and Signal Processing* **19**(4), 865-880 (2005).
15. H. Sarmadi, A. Entezami, B. Saeedi Razavi, and K.-V. Yuen, "Ensemble learning-based structural health monitoring by Mahalanobis distance metrics". *Structural Control and Health Monitoring* **28**(2), e2663 (2021).
16. A. Entezami, H. Sarmadi, M. Salar, C. De Michele, and A. Nadir Arslan, "A novel data-driven method for structural health monitoring under ambient vibration and high dimensional features by robust multidimensional scaling". *Structural Health Monitoring*, (2021).
17. A. Entezami, H. Sarmadi, B. Behkamal, and S. Mariani, "Health Monitoring of Large-Scale Civil Structures: An Approach Based on Data Partitioning and Classical Multidimensional Scaling". *Sensors* **21**(5), 1646 (2021).
18. H. Sarmadi, A. Entezami, B. Behkamal, and C. De Michele, "Partially online damage detection using long-term modal data under severe environmental effects by unsupervised feature selection and local metric learning". *Journal of Civil Structural Health Monitoring*, (2022).
19. D. Cheng, Q. Zhu, J. Huang, Q. Wu, and L. Yang, "Clustering with Local Density Peaks-Based Minimum Spanning Tree". *IEEE Transactions on Knowledge and Data Engineering* **33**(2), 374-387 (2021).
20. Q. Zhu, J. Feng, and J. Huang, "Natural neighbor: A self-adaptive neighborhood method without parameter K". *Pattern Recognition Letters* **80**, 30-36 (2016).
21. R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(2), 411-423 (2001).
22. E. Reynders and G. De Roeck, *Continuous Vibration Monitoring and Progressive Damage Testing on the Z24 Bridge*, in *Encyclopedia of Structural Health Monitoring*. 2009, John Wiley & Sons, Inc.: Chichester, United Kingdom.