Investigation of Machine Learning Methods for Structural Safety Assessment under Variability in Data: Comparative Studies and New Approaches

Hassan Sarmadi[1]

Abstract: Due to the importance of civil structures and infrastructures, structural safety assessment or structural health monitoring has become a basic necessity for every society. Recent developments of sensing and data acquisition systems enable civil engineers to exploit machine learning methods based on data-driven strategies for structural safety assessment and damage detection. However, the choice of an appropriate machine learning method may be problematic, particularly under some challenging issues such as the negative effects of environmental and/or operational variability (EOV), and the necessity of estimating some influential unknown elements of parametric machine learning methods called hyperparameters. Accordingly, this article focuses on three main aspects: (1) comparing various machine learning methods, (2) developing semiparametric algorithms, and (3) proposing automated algorithms for hyperparameter optimization of semiparametric and parametric machine learning methods. An innovative automated output-only approach is proposed to qualitatively and relatively predict the levels of EOV in terms of strong or weak variability. The main contributions of this article include comparing various machine learning methods, which will enable civil engineers to choose the most appropriate technique, and proposing automated approaches to hyperparameter optimization and variability level prediction. Dynamic and statistical features extracted from measured vibration data of two full-scale bridges were considered to perform the comparative studies and investigate the proposed methods. The results demonstrated that the semiparametric methods provide the best performance when their unknown parameters are determined appropriately. **DOI: 10.1061/(ASCE)CF.1943-5509.0001664.**

Author keywords: Structural health monitoring; Damage detection; Machine learning; Hyperparameter optimization; Environmental and operational variability; Bridges.

Introduction

Civil engineering structures are prominent structural systems that play crucial roles in social life, transportation networks, business, economic, and so forth. These systems are subjected to permanent dead and live loads along with temporary external loads caused by natural or artificial hazards. Some of these structures were designed and constructed several years ago, in which case aging and material deterioration are other hazardous factors that threaten their safety and integrity. To prevent these structures from damage, failure, and even collapse, structural safety assessment or structural health monitoring (SHM) is an important and real necessity for most civil engineering structures (Agdas et al. 2016; Jeong et al. 2020; Liu et al. 2020; Sarmadi and Yuen 2021; Seo et al. 2016; Sun et al. 2020; Yekrangnia and Mobarake 2016).

Due to recent advances in sensing and data acquisition systems, the use of raw measured data and data-driven methods has become increasingly appealing to researchers and civil engineers. The major benefit of these methods is the ability to avoid constructing an elaborate finite-element model and implementing model updating, particularly for large-scale structures. Most data-driven methods used in SHM consist of two main steps, feature extraction and feature classification (Farrar and Worden 2013). Feature extraction is intended to extract meaningful information (i.e., damage-sensitive features) from raw measured data such as acceleration time histories, images (Jeong et al. 2020), and videos (Yang et al. 2017). Feature classification is a process for decision-making via machine learning algorithms (Flah et al. 2020). The main objective of this process is to utilize features extracted from raw measured data to make a decision about the current status of a structure in terms of indicating the occurrence of damage or declaring a safe or normal condition.

In most cases of SHM applications, machine learning algorithms are often classified as supervised learning and unsupervised learning (Farrar and Worden 2013). Because supervised learning requires full labeled data from both the normal and damaged conditions for the learning process, unsupervised learning is more beneficial to use in SHM applications. The main reason for this selection is that it is not reasonable and economical to impose intentional damage patterns in an effort to provide information about the damaged state of a valuable and important civil structure (Sarmadi and Entezami 2021). The implementation of a SHM strategy by unsupervised learning methods is carried out in training (baseline) and monitoring (inspection) phases (Sarmadi and Karamodin 2020). During the training period, a statistical model is learned using the features of the normal condition (i.e., training data). The outputs of this model then are used to estimate an alarming threshold for decision-making. In the monitoring phase, the same feature extraction method is applied to extract features of the current state (i.e., test data) and

[1]Researcher, Dept. of Civil Engineering, Faculty of Engineering, Ferdowsi Univ. of Mashhad, P.O. Box: 91779-48974, Mashhad, Iran; Head of Research and Development, Ideh Pardazan Etebar Sazeh Fanavar Pooya (IPESFP) Company, 29th Reza St., Reza Blvd., P.O. Box: 91767-68540, Mashhad, Iran. ORCID: https://orcid.org/0000-0002-2985-091X. Email: hassan.sarmadi@mail.um.ac.ir; sarmadi.ipesfp@gmail.com

Note. This manuscript was submitted on June 3, 2021; approved on July 28, 2021; published online on September 22, 2021. Discussion period open until February 22, 2022; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Performance of Constructed Facilities***, © ASCE, ISSN 0887-3828.

feed them into the learned model. Any deviation of the outputs of the model regarding the test samples from the model of interest (i.e., the alarming threshold) is indicative of damage occurrence (Sarmadi and Entezami 2021).

Unsupervised learning methods for feature classification can be divided into nonparametric, semiparametric, and parametric approaches. A nonparametric method does not need to estimate any unknown parameter. This is the greatest advantage of this kind of unsupervised learning method: it is more computationally efficient than the other techniques. Statistical distance measures (Entezami et al. 2020b; Sarmadi et al. 2021a), singular value decomposition (SVD) (Figueiredo et al. 2011), and robust multidimensional scaling (Entezami et al. 2021) are some effective nonparametric unsupervised learning methods. A parametric method is one that depends on unknown parameters that affect its performance. Partition-based clustering techniques such as k-medoids (Sarmadi et al. 2021b), fuzzy c-means (Yu and Zhu 2017), Gaussian mixture model (GMM) (Figueiredo et al. 2019), and self-organizing map (Avci and Abdeljaber 2016), and various artificial neural networks (ANNs) (Bagchi et al. 2010; Chang et al. 2018; Weinstein et al. 2018), are popular parametric unsupervised learning methods. In these approaches, the numbers of clusters, layers, and neurons are unknown parameters that should be determined properly. A semiparametric method is a combination of nonparametric and parametric algorithms. In other words, this method is developed in a nonparametric manner, and a parametric scheme is added to improve its performance. Some recent examples of semiparametric methods are adaptive Mahalanobis squared distance (MSD) and one-class k-nearest neighbor (Sarmadi and Karamodin 2020), MSD and GMM (Daneshvar et al. 2021), two distance measures, and robust principal component analysis (Gharibnezhad et al. 2015).

Despite various studies of SHM of civil structures, some major challenges still are open problems that should be dealt with properly. One is related to the negative effects of environmental and/or operational variability (EOV) (Farrar and Worden 2013) that may cause false alarm (false positive or Type 1) and false detection (false negative or Type 2) errors (Sarmadi et al. 2021a). The EOV conditions are major challenges in SHM because they may lead to changes in inherent physical properties such as mass and stiffness, as well as the dynamic characteristics of civil structures, which may be interpreted mistakenly as damage and cause Type 1 errors and economic losses. On the other hand, these conditions may have magnitudes of changes that are equal to or larger than those of damage. In this case, one cannot accurately indicate the occurrence of damage, particularly in minor damage scenarios, leading to safety losses caused by Type 2 errors. In the context of SHM, these erroneous situations are known as outlier masking problems. Therefore, it is vital to consider the EOV conditions and attempt to remove the outlier masking effects from features extracted from measured data.

The other major challenge is a methodological issue regarding semiparametric and parametric methods, for which some unknown parameters play important roles in their performance. This issue refers to the problem of hyperparameter optimization. In machine learning, a hyperparameter is an unknown parameter the value of which is used to control the learning process. On this basis, hyperparameter optimization is a problem of choosing or tuning a set of optimal unknown parameters for a learning algorithm (Feurer and Hutter 2019; Yang and Shami 2020). In this problem, two kinds of parameters can be estimated or optimized during the learning process. The first type (i.e., model parameters) can be initialized and updated through the learning procedure (e.g., the weights of neurons in an artificial neural network). The second type (i.e., hyperparameters) designs the configuration of a machine learning model and strongly affects its performance (e.g., the numbers of clusters and neurons). Because this type of unknown parameter may not be estimated directly from the data learning process, it is necessary to choose hyperparameters before learning the model of interest (Yang and Shami 2020). This process can be carried out by manual tuning or gradient-based optimization techniques. Manual tuning (e.g., grid and random searches) is the simplest method for tuning unknown parameters of a machine learning model by considering a relatively large number of sample parameters and finding the most appropriate sample yielding the best performance of the model. Gradient-based optimization is based on defining an objective (mathematical) function and optimizing it to estimate some unknown parameters (Feurer and Hutter 2019; Yang and Shami 2020). Although both hyperparameter optimization techniques have their advantages and disadvantages, the manual tuning method may prevail against gradient-based optimization, particularly when there are a few hyperparameters. Furthermore, it may be difficult to define an objective function for a special issue (e.g., dealing with the outlier masking problem). Finally, the last challenge is related to the necessity of conducting a comprehensive comparative study of different machine learning methods, feature types, and their dimensions in terms of high-dimensionality and low-dimensionality, and threshold estimation techniques.

Accordingly, this article investigated various methods of feature extraction and feature classification, particularly nonparametric (i.e., MSD and SVD), semiparametric, and parametric (fuzzy c-means and GMM) algorithms, and proposed automated approaches to tuning some important hyperparameters. The combinations of MSD and SVD with ANNs made two different semiparametric methods, so that the neuron sizes of the hidden layers of the ANNs were the main hyperparameters. The parametric methods were developed from the fuzzy c-means and GMM clustering algorithms, for which the cluster numbers should be determined properly. The main reason for proposing the hyperparameter optimization methods was to deal with the outlier masking problem caused by the EOV conditions. For this, this article exploited the median absolute deviation (MAD), which is a robust statistical measure for computing dispersion and variability in data. The main innovation of these methods is the development of automated strategies based on the MAD criterion. The great advantage of these approaches is their ability to tune any unknown parameters without defining any objective function and solving it by computational techniques. Due to the great importance of the EOV conditions, on the other hand, an innovative output-only approach based on the MAD was proposed to predict the relative levels of EOV in terms of strong versus weak variability. Development of an automated output-only algorithm without any EOV data was the main innovation of this approach. Dynamic and statistical features of two full-scale bridges under actual EOV conditions were utilized to compare and validate the methods presented in this article. Results showed that the proposed semiparametric methods provide the best performance when their hyperparameters are determined properly. Moreover, the proposed output-only approach is a useful tool for predicting the levels of EOV conditions.

Nonparametric Machine Learning Methods

MSD-Based Nonparametric Algorithm

The MSD is a statistical distance that measures the dissimilarity between two multivariate sets by considering the correlation among variables. In other words, it is a metric for measuring the distance of a point from the center of a distribution in a multivariate space. The main advantages of the MSD-based machine learning method are

its nonparametric property, simplicity, and computational efficiency. Suppose that $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in \mathbb{R}^{p\times n}$ is the training set containing some features extracted from measured vibration data of the normal condition in the training (baseline) phase. The key components of the machine learning model derived from the MSD are the mean vector $\boldsymbol{\mu}_{\mathbf{x}} \in \mathbb{R}^p$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{x}} \in \mathbb{R}^{p\times p}$ of the training data. These components are normally estimated by the sample mean and covariance under the Gaussianity assumption of $\mathbf{X}$. Assume that $\mathbf{z}$ is a new p-dimensional feature vector from the test set $\mathbf{Z} \in \mathbb{R}^{p\times m}$ regarding the remaining features (validation samples) of the normal condition and all features of the current state of the structure in the monitoring (inspection) phase. Therefore, the formulation of MSD for damage detection is expressed as follows:

$$\mathrm{DI}_m = (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{x}})^T \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{z} - \boldsymbol{\mu}_{\mathbf{x}}) \tag{1}$$

where DI_m denotes a damage index (DI) related to the MSD technique. If the current state of the structure suffered from damage, the new feature vector $\mathbf{z}$ is concerned with the damaged condition. In this case, the vector will be farther from the mean of the normal features implying the emergence of damage in the structure (Figueiredo et al. 2011). On the other hand, the feature vectors in $\mathbf{X}$ are also used in Eq. (1) to determine the distance values of the normal condition in the training phase. These distance quantities are used to estimate an alarming threshold for damage detection (Sarmadi and Entezami 2021). Therefore, any deviation of the distance value DI_m from the threshold limit is indicative of damage occurrence.

SVD-Based Nonparametric Algorithm

In mathematics and linear algebra, the SVD is a technique for decomposing a whole matrix into three submatrices. Despite this usual application, Ruotolo and Surace (1999) proposed a damage detection method via the fundamental principle of SVD. This method is based on determining the rank of a state matrix $\mathbf{M}$ (where the rank of a matrix is defined as the maximum number of linearly independent columns or rows in the matrix of interest) by using the singular values of the feature vectors of the training matrix and a new feature vector related to the test matrix. Simplicity, computational efficiency, and a nonparametric characteristic are the main advantages of the SVD-based machine learning method (Figueiredo et al. 2011).

Assume that the rank of the training matrix $\mathbf{X} \in \mathbb{R}^{p\times n}$ corresponds to b, which means that b feature vectors of $\mathbf{X}$ are independent. Using the new feature vector $\mathbf{z}$ of the current condition, the state matrix is defined as $\mathbf{M} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_b, \mathbf{z}]$. If the feature vector $\mathbf{z}$ comes from the normal condition of the structure, it is expected that the rank of the state matrix $\mathbf{M}$ will not change and it is equal to b. Conversely, if the feature vector $\mathbf{z}$ originates from the damaged condition, it is independent of the b feature vectors $\mathbf{x}_1 \ldots \mathbf{x}_b$ and the rank will be $b + 1$. Rather than using the concept of the variation in the rank of the state matrix, which may be problematic when the outlier masking problem is a key issue, a nonparametric algorithm is presented to define a damage index (DI) for damage detection. This algorithm utilizes the residual (discrepancy) between the vectors of the singular values related to the state matrix $\mathbf{M}$ and the training matrix $\mathbf{X}$. Accordingly, the DI or distance value of the SVD-based machine learning method is expressed as $\mathrm{DI}_s = \|\boldsymbol{\delta}_{\mathbf{X}} - \boldsymbol{\delta}_{\mathbf{M}}\|_2$, where $\boldsymbol{\delta}_{\mathbf{X}}$ and $\boldsymbol{\delta}_{\mathbf{M}}$ are the vectors of the singular values of the training and state matrices, respectively, and $\|\cdot\|_2$ indicates the Euclidean or l_2-norm. For each feature vector of the current state, it is possible to obtain a DI_s quantity, in which case the deviation of this quantity from an alarming threshold implies the occurrence of damage. The threshold is determined using only the DI_s values obtained from the training phase.

Semiparametric Machine Learning Methods

Although the MSD and SVD methods present nonparametric algorithms for damage detection, the outlier masking problem may seriously affect their performance and cause considerable Type 1 or Type 2 errors under strong variability. To deal with this issue, the proposed semiparametric methods ANN-MSD and ANN-SVD are considered to initially remove potential EOV conditions from the features of the training and test matrices ($\mathbf{X}$ and $\mathbf{Z}$) via ANNs and then detect damage by the MSD and SVD. For the sake of convenience, Fig. 1 is a flowchart of the decision-making procedure using either nonparametric or semiparametric approaches. In Figs. 1(a and b), Steps 1–3 and 4–6 are related to the MSD and SVD methods in the training and monitoring phases as discussed in the sections "MSD-Based Nonparametric Algorithm" and "SVD-Based Nonparametric Algorithm." Moreover, Fig. 1(c) shows the initial

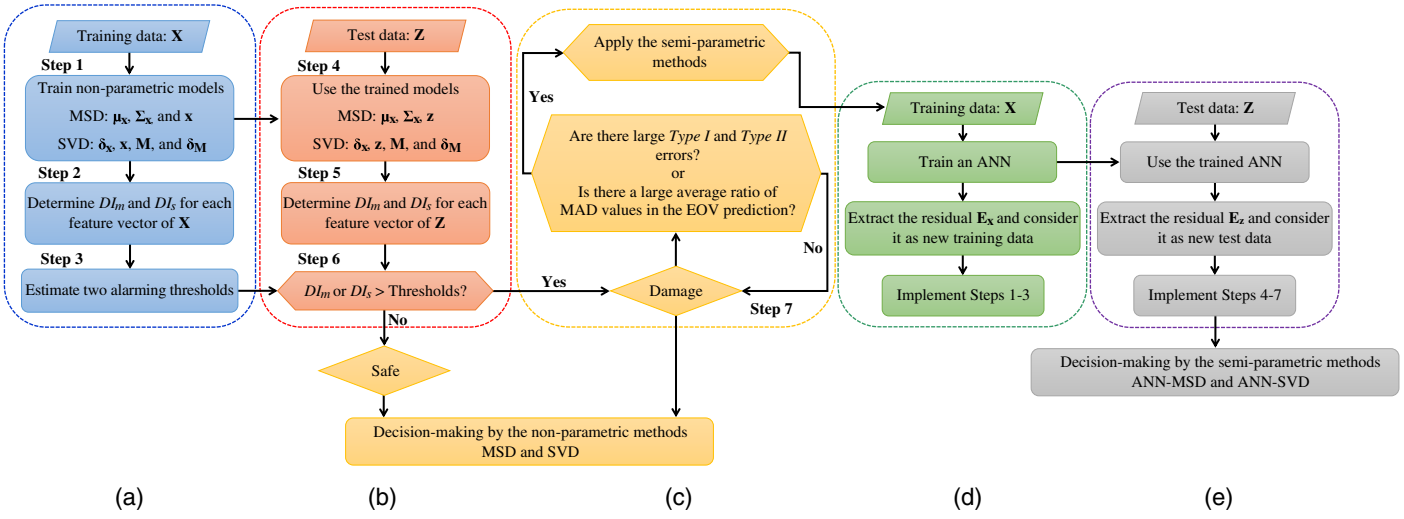

Fig. 1. Flowchart of damage detection by the nonparametric and semiparametric methods: (a) training phase and nonparametric methods; (b) monitoring phase and nonparametric methods; (c) decision-making process; (d) training phase and semiparametric methods; and (e) monitoring phase and semiparametric methods.

decision-making process based on the outputs of the nonparametric methods and the decision-making criterions; that is, the amounts of Type 1 and Type 2 errors or the outputs of the EOV level prediction (section "Prediction of EOV Conditions"). In the case of low errors or small MAD values, one can utilize the MSD and SVD techniques; otherwise, one should apply the ANN-MSD and ANN-SVD methods. Generally, it is recommended to apply the semiparametric methods unless the criterion outputs are inconsiderable.

Because the underlying concept of the proposed semiparametric methods is to use an effective ANN for addressing the outlier masking, this article exploited an autoassociative neural network under a feed-forward configuration for reconstructing the input data (Kramer 1992). The autoassociative neural network is an ANN that contains an input layer; three hidden layers called mapping, bottleneck, and demapping; and an output layer. The great benefit of this neural network is its unsupervised learning behavior and its great capability to filter out noise, outliers, and any type of variations in data such as the EOV (Figueiredo et al. 2011; Kramer 1992). Because the number of layers required for configuring the neural network is known (i.e., five layers), the only unknown parameter is the neuron sizes of the mapping, bottleneck, and demapping layers. The bottleneck layer of the autoassociative neural network should have smaller neurons than the other hidden layers (Kramer 1992). Furthermore, the neuron sizes of the mapping and demapping layers are often similar. Therefore, the process of hyperparameter optimization relates to determining the numbers of neurons of mapping and bottleneck layers (section "Automated Selection of Optimal Neurons").

To remove the EOV conditions, the training data is firstly applied to learn an auto-associative neural network. Subsequently, the initial features in the training and test matrices $\mathbf{X}$ and $\mathbf{Z}$ are fed into the learned neural network as the inputs to obtain the outputs $\mathbf{X}^*$ and $\mathbf{Z}^*$ that have the same sizes as the inputs. Finally, the residuals or discrepancies between the inputs and outputs are calculated as $\mathbf{E}_\mathbf{x} = \mathbf{X} - \mathbf{X}^*$ and $\mathbf{E}_\mathbf{z} = \mathbf{Z} - \mathbf{Z}^*$ [Figs. 1(d and e)]. These matrices are then incorporated into the MSD and SVD algorithms instead of the matrices $\mathbf{X}$ and $\mathbf{Z}$.

Parametric Machine Learning Methods

Clustering is an unsupervised learning method that aims to divide or segment similar data points with small distances into clusters. Based on this general concept, it is feasible to utilize various clustering methods based on the partition-, density-, graph-based, and hybrid algorithms (Aggarwal and Reddy 2016). Among them, partition-based clustering techniques such as *k*-means, *k*-medoids, fuzzy *c*-means, and GMM are more well-known to SHM applications. In this article, the fuzzy *c*-means and GMM were selected as the main parametric machine learning methods.

Clustering in SHM consists of two main steps. First, one needs to determine the number of clusters or components by using the training data $\mathbf{X}$. The main objective of this step is to obtain the clustering outputs depending upon the type of clustering algorithm and its structure. Second, it is necessary to define a DI (in most cases, based on the Euclidean or Mahalanobis distances) for calculating the dissimilarity or distance of each feature vector in $\mathbf{X}$ from the clustering outputs. The second step is also implemented by using the feature vector of $\mathbf{Z}$ in the inspection phase (Sarmadi et al. 2021b).

Fuzzy c-Means Clustering

Fuzzy *c*-means clustering is an unsupervised learning method based on the concept of fuzzy logic that allows one sample of data to be put in two or more clusters based on an objective function. Give the feature vector $\mathbf{x}$ from the matrix $\mathbf{X} \in \Re^{p\times n}$ and clusters $\mathbf{v}_1 \dots \mathbf{v}_c$, fuzzy *c*-means is based on the minimization of the following objective function (Aggarwal and Reddy 2016):

$$J_f = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij}^{\beta}\|\mathbf{x}_i - \mathbf{v}_j\|_2^2 \tag{2}$$

where $1 < \beta < \infty$ = weighting fuzziness; and u_{ij} = degree of membership value of *i*th data in *j*th cluster, which varies from 0 to 1. The algorithm of fuzzy clustering is based on iteratively optimizing the objective function to update the membership u_{ij} and the *j*th cluster as (Aggarwal and Reddy 2016)

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|\mathbf{x}_i - \mathbf{v}_j\|_2^2}{\|\mathbf{x}_i - \mathbf{v}_k\|_2^2}\right)^{2/(\beta-1)}} \tag{3}$$

$$\mathbf{v}_j = \frac{\sum_{i=1}^{n} u_{ij}^{\beta}\mathbf{x}_i}{\sum_{i=1}^{n} u_{ij}^{\beta}} \tag{4}$$

The membership u_{ij} should satisfy three main conditions: (1) the variations should be between 0 and 1 (i.e., $u_{ij} \in [0\ 1]$, $\forall\ i, j$), (2) the membership values for each data sample should sum to 1 (i.e., $\sum_{j=1}^{c} u_{ij} = 1,\ \forall\ i$), and (3) the variation of the sum of all the membership values in a cluster should be between zero and n (i.e., $0 < \sum_{j=1}^{c} u_{ij} < n,\ \forall\ n$). The termination of the iterative algorithm in fuzzy *c*-means clustering occurs when the maximum difference between two successive memberships is smaller than a termination step varying between 0 and 1. Having determined the cluster number *c*, the clustering method gives the clusters $\mathbf{v}_1 \dots \mathbf{v}_c$. These vectors are considered to determine a DI_f value of the feature vector $\mathbf{z}$ as follows:

$$\mathrm{DI}_f = \min(\|\mathbf{z} - \mathbf{v}_1\|_2, \dots, \|\mathbf{z} - \mathbf{v}_c\|_2) \tag{5}$$

The same distance calculation is carried out by using the feature samples of the training data $\mathbf{X}$ in the training phase, instead of $\mathbf{z}$, to obtain n values of DI_f. These distance values are utilized to estimate an alarming threshold. In this regard, any deviation of DI_f of the feature vector $\mathbf{z}$ from this threshold is representative of damage; otherwise, the structure is safe in its current state.

Gaussian Mixture Model

The GMM is an unsupervised learning method under probability theory that defines a probability model for clustering under the assumption that all data samples are generated from a mixture of *r* component Gaussian distributions (Aggarwal and Reddy 2016). In this method, the underlying assumption is that all data samples to be clustered are drawn from some proper components; hence, the problem is to estimate the unknown parameters of each component to fit the data properly. A GMM as a weighted sum of *r* component Gaussian densities is formulated as $\rho(\mathbf{x}|\lambda) = \sum_{i=1}^{r}\omega_i g(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, where $\mathbf{x}$ is one of the feature vectors of the training matrix $\mathbf{X}$, ω_i is the *i*th mixture weight, and $g(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ is the Gaussian density of the *i*th component. The mixture weights should satisfy a constraint that $\sum_{i=1}^{r}\omega_i = 1$. Each component density is representative of a Gaussian function in the following form (Aggarwal and Reddy 2016):

$$g(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}_i|^{n/2}}\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right) \tag{6}$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ = mean vector and covariance matrix of *i*th component, respectively. Apart from the number of components *r* (a hyperparameter), the GMM includes some unknown model

parameters such as the mean vector, the covariance matrix, and the mixture weights, all of which can be tuned during the clustering process. In most cases, the maximum-likelihood estimation based on the expectation-maximization (EM) algorithm is applied to estimate these parameters (Daneshvar et al. 2021; McLachlan and Krishnan 2007). For early damage detection via the GMM, it is only necessary to use the estimated mean vectors and covariance matrices of all r components and determine a distance value for the feature sample $\mathbf{z}$ as

$$\mathrm{DI}_g = \min(d_1, \ldots, d_i, \ldots, d_r) \tag{7}$$

where $d_i = (\mathbf{z} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{z} - \boldsymbol{\mu}_i)$. The same distance calculation can be performed by using the feature vectors of the training matrix $\mathbf{X}$ in the training phase, instead of $\mathbf{z}$, to determine n values of DI_g. These distance values are incorporated to estimate an alarming threshold. Accordingly, any deviation of DI_g of the feature vector $\mathbf{z}$ from this threshold is representative of damage; otherwise, the structure is safe in its current state.

Proposed Hyperparameter Optimization Methods

The selection of hyperparameters is crucial to the semiparametric and parametric methods due to the direct impact of such parameters on the performance of these methods. Furthermore, this selection depends on the main challenge and/or limitation of the problem under study. As mentioned previously, one of the major demanding issues in SHM is the problem of outlier masking caused by the EOV conditions and its negative effects (i.e., Type 1 and 2 errors). Accordingly, a large rate of any error implies poor performance of each of the machine learning methods, which may lead to economic or safety losses. Therefore, this article developed automated approaches based on the MAD criterion for choosing the hyperparameters of the semiparametric and parametric techniques by dealing with the outlier masking problem. In ANN-MSD and ANN-SVD, the hyperparameters include the neuron sizes of the mapping (demapping) and bottleneck layers. Moreover, the cluster numbers c and r are the main hyperparameters of the fuzzy c-means and GMM techniques.

Median Absolute Deviation

In statistics, the MAD is a robust indicator for measuring variability in a univariate set of sampling data and detecting outliers (Dodge 2008). Although the variance and standard deviation of a set of random samples also measure the dispersion and variability, the major merit of the MAD is its suitability for measuring the variability of non-Gaussian data. Moreover, this measure is more resilient to outliers than the variance and standard deviation (Leys et al. 2013). Let $y_1, \ldots, y_q$ be a set of q data samples, which are equivalent to the DI values of the semiparametric and parametric methods. On this basis, the MAD criterion (η) is defined as (Dodge 2008)

$$\eta = \frac{1}{q} \sum_{i=1}^{q} |y_i - \xi| \tag{8}$$

where ξ = median of $y_1, \ldots, y_q$; and $|\cdot|$ is the absolute operator. This value is a measure of central tendency similar to the mean, but is more resilient to outliers.

Automated Selection of Optimal Neurons

The automated process of choosing the optimal neurons of the hidden layers of the autoassociative neural network is based on three main steps: (1) examining a relatively large number of sample neurons, (2) filtering out the EOV conditions from the initial training samples, and (3) computing the MAD values of the DI quantities obtained from the MSD and SVD methods by using the residual matrix $\mathbf{E}_\mathbf{x}$. The criterion for hyperparameter optimization in this issue is to choose the neuron sizes among all sample neurons with the minimum MAD values. This means that the selected neurons enable the neural network to provide the smallest rate of variability in the DI values and filter out the EOV conditions. Because the hidden layers of the autoassociative neural network can be configured symmetrically, the mapping and demapping layers have the same neurons. Furthermore, the number of neurons in the bottleneck layer (l_b) should be smaller than the number of neurons in the mapping and demapping layers (l_m) (Kramer 1992). Because a relatively large number of sample neurons are examined, it is essential to check the overfitting problem. For this issue, one can utilize the equations $2l_m < p(n - l_b)/(p + l_b + 1)$ and $2l_m < n$ after selecting the suitable neurons (Kramer 1992). In these equations, l_m and l_b denote the optimal neuron sizes of the mapping (demapping) and bottleneck layers, respectively. If the occurrence of this problem is inevitable, one should select the next minimum MAD value. This procedure continues until the concern about the overfitting problem is dealt with.

Algorithm 1. Automated Selection of Optimal Neurons of Hidden Layers

Inputs: Training data $\mathbf{X}$, and the sample neuron numbers l_{m_0} and l_{b_0} (where $l_{m_0} > l_{b_0}$)
Iterative process
For $i = 1$ **to** l_{m_0}
  For $j = 1$ **to** l_{b_0}
    1. Train an autoassociative neural network using the ith and jth neurons of the mapping and bottleneck layers.
    2. Extract the network output $\mathbf{X}^*$.
    3. Compute the residual matrix $\mathbf{E}_\mathbf{x} = \mathbf{X} - \mathbf{X}^*$.
    4. Set the residual matrix $\mathbf{E}_\mathbf{x}$ as new training data in each of the nonparametric methods (i.e., MSD and SVD)
    5. Determine n values of DI_m and DI_s using all feature vectors of $\mathbf{E}_\mathbf{x}$.
    6. Compute the MAD values of the n-dimensional DI quantities of the ith and jth sample neurons.
    7. Store this value in the ith row and the jth column of a matrix.
  End for
End for
8. Choose the minimum MAD value in the rows and columns of the stored matrix obtained from Step 7.
9. Check the overfitting problem via $2l_m < [p(n - l_b)]/(p + l_b + 1)$ and $2l_m < n$. If the overfitting problem is inevitable, return to Step 8 and select the next minimum MAD value until the concern about the overfitting problem is dealt with.
10. The row and column numbers of the stored matrix regarding the minimum MAD value are indicative of the optimal neuron sizes of the mapping or demapping (l_m) and bottleneck (l_b) layers, respectively.

Outputs: The optimal neuron numbers l_m and l_b

For simplicity, Algorithm 1 presents the pseudocode of the main steps of the automated approach to selecting the optimal neurons of the hidden layers. The process of choosing the neuron size is carried out under an automated iterative approach. First, one needs to assign the sample neuron numbers for the mapping or demapping (l_{m_0}) and bottleneck (l_{b_0}) layers so that $l_{m_0} > l_{b_0}$ and the training matrix $\mathbf{X}$ as the inputs. Subsequently, the iterative procedure is

started with the ith neuron of the mapping and demapping layers and jth neuron of the bottleneck layer, where $i = 1, 2, \ldots, l_{m_0}$ and $j = 1, 2, \ldots, l_{b_0}$. Using the training data $\mathbf{X}$ and the ith and jth neurons, an autoassociative neural network is trained to obtain the output data $\mathbf{X}^*$ and extract the residual matrix $\mathbf{E_x}$. This matrix subsequently is applied to the MSD and SVD methods to calculate their DI values (i.e., DI_m and DI_s) for all feature vectors of $\mathbf{E_x}$. In the last step of the iterative process, it is necessary to compute the MAD values of these DI quantities and store them in an $l_{m_0} \times l_{b_0}$ matrix. After ensuring that no overfitting problem occurs, the row and column numbers of the stored matrix regarding the minimum MAD value are chosen as the optimal neurons of the mapping (demapping) and bottleneck layers, respectively.

Automated Selection of Optimal Cluster Numbers

Similar to the previous hyperparameter optimization strategy, the process of choosing the proper number of clusters is based on the main three steps: (1) examining a relatively large number of sample clusters, (2) dividing the training samples into the predefined sample clusters, and (3) calculating the MAD values of the DI amounts of the fuzzy c-means and GMM methods. The main criterion for the hyperparameter optimization of the clustering techniques is to find the sample cluster with the minimum MAD value. Similarly, the proposed strategy for the optimal cluster selection develops an automated iterative approach by evaluating the DI values of the clustering methods to remove or reduce the negative effects of the EOV conditions. The great merit of this approach is its high suitability and applicability to all partition-based clustering techniques.

Algorithm 2. Automated Selection of Optimal Clusters

Inputs: Training data $\mathbf{X}$ and the sample cluster number h

Iterative process

For $i = 1$ **to** h

1. Divide the training data into i clusters using the fuzzy c-means and GMM clustering algorithms.
2. Extract the clustering outputs (i.e., the clusters $\mathbf{v}_1 \ldots \mathbf{v}_i$ associated with the fuzzy c-means clustering, and the mean vectors $\boldsymbol{\mu}_1 \ldots \boldsymbol{\mu}_i$ and covariance matrices $\boldsymbol{\Sigma}_1 \ldots \boldsymbol{\Sigma}_i$ for the GMM).
3. Compute n-dimensional DI values for each of the clustering methods by using all feature vectors of $\mathbf{X}$.
4. Calculate the MAD value of the obtained DI quantities related to the ith cluster.
5. Store this value in a vector for each of the clustering methods.

End for

6. Choose the minimum MAD value of the stored vector obtained from Step 5.
7. The optimal cluster number h_{opt} for each of the clustering methods is the number of elements of the stored vector with the minimum MAD value.

Output: The optimal cluster number h_{opt}, which can be each of the numbers c and r.

For the sake of convenience, Algorithm 2 describes the main steps of the proposed automated approach. In this algorithm, the inputs are the training matrix $\mathbf{X}$ and a sample cluster number (h). In the iterative process, one needs to divide the training data into i clusters, where $i = 1, 2, \ldots, h$. Subsequently, the clustering outputs (i.e., the clusters $\mathbf{v}_1 \ldots \mathbf{v}_i$ associated with the fuzzy c-means as well as the mean vectors $\boldsymbol{\mu}_1 \ldots \boldsymbol{\mu}_i$ and covariance matrices $\boldsymbol{\Sigma}_1 \ldots \boldsymbol{\Sigma}_i$ for the GMM) are extracted to calculate the n-dimensional DI values of the ith sample cluster for each of the clustering methods. Using these quantities, a MAD value is computed and stored in a vector (i.e., each clustering algorithm has a vector of the MAD values). In the following, the number of elements of the h-dimensional stored vector with the minimum MAD value is set as the optimal cluster number (h_{opt}). Note that, h_{opt} is equivalent to each of the cluster numbers c and r associated with the fuzzy c-means and GMM clustering methods.

Prediction of EOV Conditions

In real-world SHM applications, the existence of EOV is inevitable. Before dealing with this challenge, it is very appropriate to relatively and qualitatively predict their levels, which may assist civil engineers in realizing the variations in data (features) in terms of strong or weak variability. Strong variability means that the vibration responses or features are highly sensitive to EOV, whereas weak variability refers to low sensitivity to these conditions. For this purpose, it is possible to utilize input–output and output-only methods. The first approach is usable when the EOV data (e.g., temperature) are available and measurable, whereas the second approach is suitable for cases in which such data are unavailable. The simplest input–output approach is to measure a correlation between the damage-sensitive features and EOV data (e.g., the correlation between modal frequencies and temperature), in which case a high correlation is indicative of strong variability. For an output-only approach, because the EOV data are not available, the level prediction is not a trivial process. Therefore, one should utilize a more rigorous strategy.

Having considered the available data (i.e., the features of the normal condition), the proposed method develops a nonparametric algorithm with the aid of the MAD criterion. The underlying steps of this approach include (1) randomly selecting training samples from all available features of the normal condition, (2) constructing a new training matrix using the selected training samples, (3) converting the new matrix into a vector, and (4) measuring the MAD value of this vector. All these steps are implemented in two stages using the initial (nonnormalized) and normalized features. The major difference between them is that, in contrast to the nonnormalized features, the EOV conditions do not exist in the normalized ones. The main objective is to compare the distance between the MAD values obtained from the initial and normalized features in two stages. In the first stage, the aforementioned steps are performed only on the initial features of the normal condition. In the second stage, one attempts to remove any EOV from the initial data and obtain normalized features using a nonparametric covariance-based technique (Deraemaeker et al. 2008). Although the second stage of this process can be implemented by an ANN (e.g., the autoassociative neural network used in the semiparametric methods) to filter out the EOV conditions, it is preferable to applying a nonparametric approach to avoid the difficulty of a hyperparameter optimization strategy.

Suppose that $\tilde{\mathbf{X}} \in \mathbb{R}^{p \times n}$ is the training matrix generated by randomly choosing the feature samples of the normal condition. Subsequently, its covariance matrix $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{p \times p}$ is decomposed into two matrices $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{S}}$ by the well-known SVD technique; that is, $\tilde{\boldsymbol{\Sigma}} = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{U}}^T$ and $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T = \mathbf{I}$, where, the diagonal matrix $\tilde{\mathbf{S}}$ consists of the singular values of the covariance matrix. This matrix is then spilt into two submatrices $\tilde{\mathbf{S}}_1 = \mathrm{diag}(\tilde{\sigma}_1 \ldots \tilde{\sigma}_s)$, which is a diagonal matrix of the first s singular values ranked in descending order, and $\tilde{\mathbf{S}}_2 = \mathrm{diag}(\tilde{\sigma}_{s+1} \ldots \tilde{\sigma}_n)$. With these descriptions, the only unknown parameter is the number of singular values (s), which can be determined simply by an indicator in a nonparametric manner as follows:

$$\gamma = \frac{\sum_{j=1}^{s} \tilde{\sigma}_j}{\sum_{i=1}^{n} \tilde{\sigma}_i} \tag{9}$$

Based on Eq. (9), the number of singular values s is one in which the lowest integer γ is larger than a threshold value (e.g., 95%). By obtaining this number, the matrix $\tilde{\mathbf{U}}$ is split into the submatrices $\tilde{\mathbf{U}}_1 = [\tilde{\mathbf{u}}_1, \ldots, \tilde{\mathbf{u}}_i, \ldots, \tilde{\mathbf{u}}_s]$ and $\tilde{\mathbf{U}}_2 = [\tilde{\mathbf{u}}_{s+1}, \ldots, \tilde{\mathbf{u}}_j, \ldots, \tilde{\mathbf{u}}_n]$, where $\tilde{\mathbf{u}}_i$ and $\tilde{\mathbf{u}}_j$ are the ith and jth column vectors of the submatrices. Using all the submatrices, one can define matrices $\mathbf{\Lambda} = \tilde{\mathbf{U}}_1(\tilde{\mathbf{S}}_1)^{1/2}$ and $\mathbf{\Psi} = \tilde{\mathbf{U}}_2\tilde{\mathbf{S}}_2\tilde{\mathbf{U}}_2^T$. To remove the EOV conditions and obtain normalized features, it is necessary to calculate the residual matrix $\tilde{\mathbf{E}} = |\tilde{\mathbf{X}} - \mathbf{\Lambda}\mathbf{H}|$, where $\mathbf{H}$ is the matrix of the unobservable factors that should be estimated. In this study, this matrix was determined by the maximum likelihood approach based on Bartlett's factor score (Deraemaeker et al. 2008) as

$$\mathbf{H} = (\mathbf{\Lambda}^T\mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^T\mathbf{\Psi}^{-1}\tilde{\mathbf{X}} \tag{10}$$

After the residual matrix has been obtained, it should be converted into a vector to calculate its MAD value in the second stage of the proposed prediction method. The first and second stages are repeated R times (e.g., $R = 1{,}000$) by randomly choosing the feature samples of the normal condition. Accordingly, one can qualitatively predict the level of EOV by comparing the MAD values of the first and second stages or measuring the average ratio of these MAD quantities. In this case, a relatively large distance or average ratio is representative of strong variability or high sensitivity to the EOV conditions, and vice versa. For convenience, Algorithm 3 lists the main steps of the pseudocode for the process of the EOV level prediction.

Algorithm 3. Prediction of EOV Conditions

Inputs: All feature samples of the normal condition, and a repetition number R (e.g., 1,000)

Stage 1

For $i = 1$ **to** R

1. Generate a training matrix $\tilde{\mathbf{X}}$ by randomly choosing the feature samples of the normal condition and store this matrix for the second stage.
2. Convert the matrix $\tilde{\mathbf{X}}$ into a vector.
3. Calculate the MAD value of this vector.
4. Store this value in a vector for the ith iteration.

End for

Stage 2

For $i = 1$ **to** R

5. Load the training matrix $\tilde{\mathbf{X}}$ from the previous stage regarding the ith iteration.
6. Estimate the covariance matrix $\tilde{\mathbf{\Sigma}}$ and decompose it into the matrices $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{S}}$, where $\tilde{\mathbf{\Sigma}} = \tilde{\mathbf{U}}\tilde{\mathbf{S}}\tilde{\mathbf{U}}^T$ and $\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T = \mathbf{I}$.
7. Determine the number of singular values of the matrix $\tilde{\mathbf{S}}$ via Eq. (9).
8. Decompose the matrices $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{U}}$ into the submatrices $\tilde{\mathbf{S}}_1$, $\tilde{\mathbf{S}}_2$, $\tilde{\mathbf{U}}_1$, and $\tilde{\mathbf{U}}_2$.
9. Compute the matrices $\mathbf{\Lambda} = \tilde{\mathbf{U}}_1(\tilde{\mathbf{S}}_1)^{1/2}$, $\mathbf{\Psi} = \tilde{\mathbf{U}}_2\tilde{\mathbf{S}}_2\tilde{\mathbf{U}}_2^T$, and $\mathbf{H} = (\mathbf{\Lambda}^T\mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}\mathbf{\Lambda}^T\mathbf{\Psi}^{-1}\tilde{\mathbf{X}}$.
10. Compute the residual matrix $\tilde{\mathbf{E}} = |\tilde{\mathbf{X}} - \mathbf{\Lambda}\mathbf{H}|$.
11. Implement Steps 2–4 using the residual matrix $\tilde{\mathbf{E}}$ instead of $\tilde{\mathbf{X}}$.

End for

12. Calculate the distance and average ratio between the MAD values of the first and second stages.

Output: The distance and average ratio of the MAD values between $\tilde{\mathbf{E}}$ and $\tilde{\mathbf{X}}$.

Performance Evaluation

In this section, the performance and effectiveness of the machine learning methods and the proposed approaches are compared and evaluated using two different features extracted from vibration responses of two full-scale structures. The first structure is the Z24 Bridge, which was described fully by Reynders and De Roeck (2009), for which modal frequencies of a few modes were used as the main dynamic features. The second structure is the Tianjin–Yonghe Bridge, fully discussed by Li et al. (2014), for which the main statistical features were extracted from the autoregressive moving average (ARMA) model (Entezami et al. 2020a). For feature classification, the 95% confidence interval under the central limit theorem (CLT) and an extreme value theorem (EVT)-based method developed by Sarmadi and Karamodin (2020) were considered to estimate two different alarming thresholds.

Z24 Bridge

The Z24 Bridge was a classical post-tensioned concrete box-girder bridge located in Switzerland. The longitudinal section, the main dimensions, and the top view of this bridge are shown in Fig. 2. To construct a new bridge with a larger side span, the Z24 Bridge was demolished in 1998. Before complete demolition, a long-term continuous SHM program was conducted to quantify the environmental variability components, including temperature, wind characteristics, and humidity, and acquire acceleration time histories from some sensors. Furthermore, realistic damage scenarios were applied gradually to the bridge in a controlled way during the month before complete demolition.

Using an operational modal analysis (i.e., stochastic subspace identification), modal frequencies of four modes were identified in a long-term monitoring scheme. The total number of measurements was 3,932 samples of modal frequencies, after eliminating some missing data. On this basis, the first 3,470 samples were related to the normal condition, and the remaining samples pertained to the damaged state of the bridge. To evaluate the effects of the environmental variability on the modal frequencies of the Z24 Bridge, Fig. 3 illustrates the MAD values of the first and second stages of the proposed output-only approach, in which the number of repetitions (R) for each stage was 1,000. The MAD values of the first stage were about 2.94, and these values significantly decrease to 0.0005–0.0006 in the second stage. Moreover, the ratio of the averages of the MAD values of the first to the second stages was 4,604. This extremely large amount and the distance between the MAD values of the first and second stages indicate the existence of strong environmental variability.

To compare the machine learning methods, training and test data sets were defined. Accordingly, 75% of all samples of the modal frequencies for the normal condition were considered to produce the training matrix $\mathbf{X} \in \mathbb{R}^{4\times 2602}$. The remaining 25% of the modal frequencies of the normal condition (Samples 2,603–3,470) served as the validation data, and all modal frequencies related to the damaged state (Samples 3,471–3,932) were applied to make the test matrix $\mathbf{Z} \in \mathbb{R}^{4\times 1330}$. Using these matrices, the process of damage detection was initially carried out using the MSD- and SVD-based methods [Figs. 4(a and b) and Figs. 4(c and d), respectively]. Horizontal dashed lines denote the alarming thresholds estimated by the EVT [Figs. 4(a and c)] and the 95% confidence interval on the basis of CLT [Figs. 4(b and d)] at a 5% significance level. In Figs. 4(a and c), the DI values of the training and validation samples concerning the normal condition of the structure (Samples 1–3,470) do not exceed the thresholds indicating any Type 1 error. This proves the great ability of the EVT, the generalized extreme value,

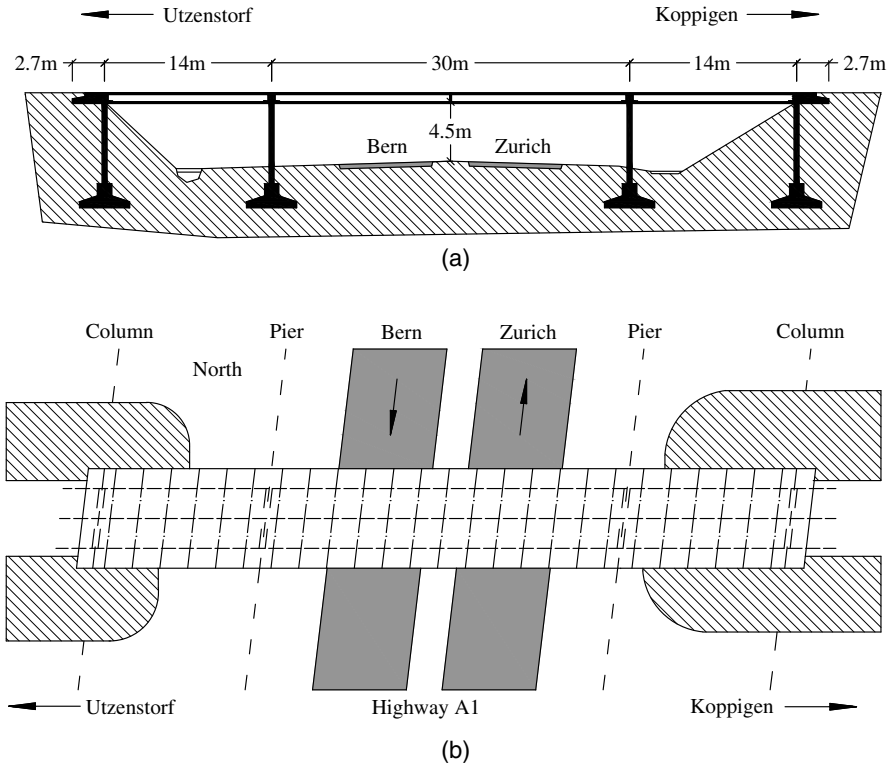

Fig. 2. (a) Longitudinal section of the Z24 Bridge; and (b) top view.

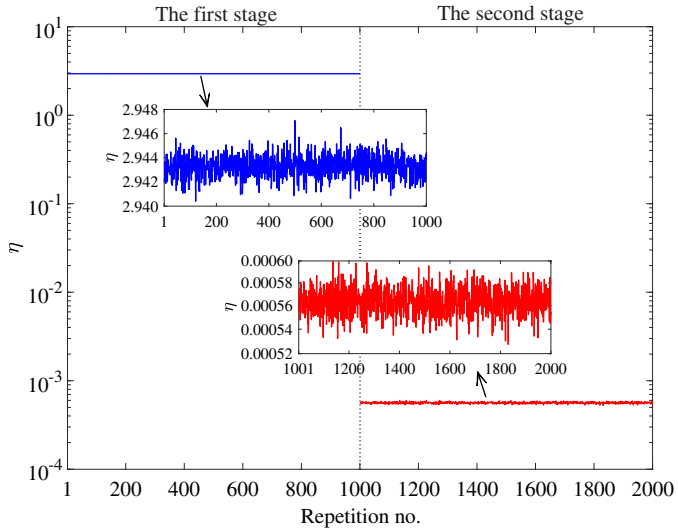

Fig. 3. Prediction of the level of environmental variability in the Z24 Bridge.

and the block maxima method to estimate a reliable threshold without any false positive. However, all DI quantities of the damaged state in Samples 3,471–3,932 mistakenly fall below the threshold lines implying considerable Type 2 errors.

In Figs. 4(b and d), many DI values for the normal condition incorrectly deviate from the thresholds and indicate large Type 1 errors. This demonstrates the poor performance of the CLT and standard confidence interval in estimating a reliable threshold limit. Moreover, these plots include serious Type 2 errors. Without considering the alarming thresholds, it is difficult to discriminate the damaged state from the normal condition. Furthermore, sudden jumps caused by the strong influence of the environmental variability on the identified modal frequencies of the normal condition are apparent in DI values of Samples 867–1,734. These conclusions imply the strong influence of environmental variability, which decreases the damage detectability of the nonparametric methods. Therefore, the nonparametric methods fail to accurately detect damage in the presence of strong environmental variations.

Due to the unreliable results of the MSD and SVD methods, the semiparametric approaches were used to evaluate their performances in detecting damage under strong environmental variations. First, the optimal neuron sizes of the mapping and bottleneck layers were determined by the proposed hyperparameter optimization technique (Algorithm 1). Fig. 5 shows the MAD values required for training the ANNs of the ANN-MSD and ANN-SVD methods, in which both sample neurons l_{m_0} and l_{b_0} were set to 30 and 10, respectively. The optimal neurons of the mapping and demapping layers (l_m) for the ANN-MSD and ANN-SVD methods were 26 and 28, respectively [Figs. 5(a and c)]. Furthermore, the neurons of the bottleneck layer (l_b) for these methods were 4 and 3, respectively [Figs. 5(b and d)]. By using these optimal neurons in the equations $2l_m < [p(n - l_b)]/(p + l_b + 1)$ and $2l_m < n$, one can ensure that the overfitting problem did not occur. By removing the environmental variations from the feature samples and determining the residual matrices $\mathbf{E_x}$ and $\mathbf{E_z}$, Fig. 6 displays the results of damage detection by the semiparametric methods, where the horizontal dashed lines are the thresholds estimated by the EVT [Figs. 6(a and c)] and the 95% confidence interval on the basis of CLT [Figs. 6(b and d)] at a 5% significance level.

In Figs. 6(a and c) there is no violation of the DI values of the training and validation samples from the thresholds gained by the EVT, except for two points in Fig. 6(c). In contrast, many exceedances of the DI quantities from the thresholds are apparent in Figs. 6(b and d) for the CLT. Regardless of the thresholds, there is a clear difference between the DI quantities of the normal and

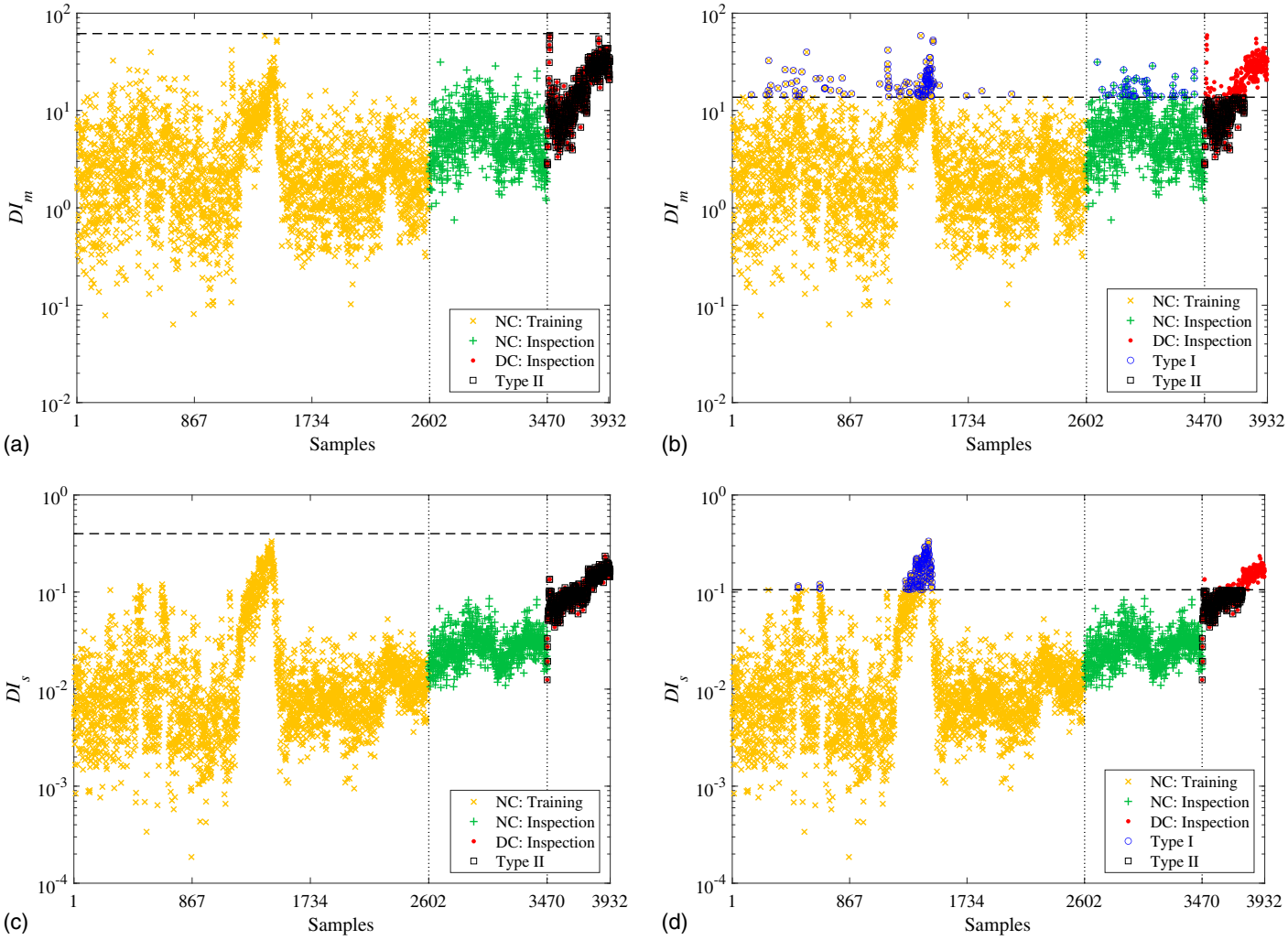

Fig. 4. Damage detection of the Z24 Bridge using the nonparametric methods: (a) MSD and EVT; (b) MSD and CLT; (c) SVD and EVT; and (d) SVD and CLT.

damaged samples, indicating good damage detectability. The results in Figs. 4 and 6 indicate that the predicted outputs of the ANN are less influenced by the environmental variability, in which case the use of the ANN greatly allows the MSD and SVD techniques to enhance the results of damage detection. On this basis, the sudden jumps in the DI values of Samples 867–1,734 decrease significantly, and the detectability of damage increases. These conclusions prove the positive effects of applying the ANN and the automated hyperparameter optimization to the outlier masking problem and removing the environmental variations from the modal frequencies.

Finally, the process of damage detection was carried out using the parametric clustering methods. Based on the proposed automated hyperparameter optimization described in Algorithm 2, Fig. 7 shows the MAD values for choosing the optimal cluster numbers c and r for the fuzzy c-means and the GMM clustering. For this problem, the sample cluster number h was set to 50. Therefore, the cluster numbers c and r were 41 and 44, respectively. Using these numbers, the results of damage detection by the parametric methods are illustrated in Fig. 8, where the horizontal dashed lines are the thresholds estimated by the EVT [Figs. 8(a and c)] and the 95% confidence interval based on the CLT [Figs. 8(b and d)] at a 5% significance level.

Once again, the EVT succeeded in estimating accurate thresholds with no Type a errors in all three clustering methods [Figs. 8(a and c)]. Except for the result of damage detection by the GMM in Fig. 8(d), the CLT was not sufficiently successful in providing an accurate threshold limit due to serious Type 1 errors [Fig. 8(b)]. The best performance in the normal condition was that of the GMM method, which entirely filtered out the environmental variations and provided DI values with low variability. Therefore, the rates of Type 1 errors in Figs. 8(c and d) are roughly similar. Concerning the damage detectability and false negative, the fuzzy c-means method was not as good as the GMM method, particularly using the EVT. Because the GMM outperformed the fuzzy c-means clustering in dealing with the outlier masking problem, this approach also had the best performance in terms of damage detectability. The values of DI_g of the normal and damaged states were distinguishable without considering the thresholds.

To summarize the performance of the machine learning methods, Table 1 lists the numbers and percentages of Type 1, Type 2, and total (misclassification) errors. The nonparametric methods (i.e., MSD and SVD) are not suitable for damage detection under strong environmental variability, owing to the substantial Type 2 and total errors even using the EVT. The best performance in terms of the smallest Type 1 and total errors was that of the proposed

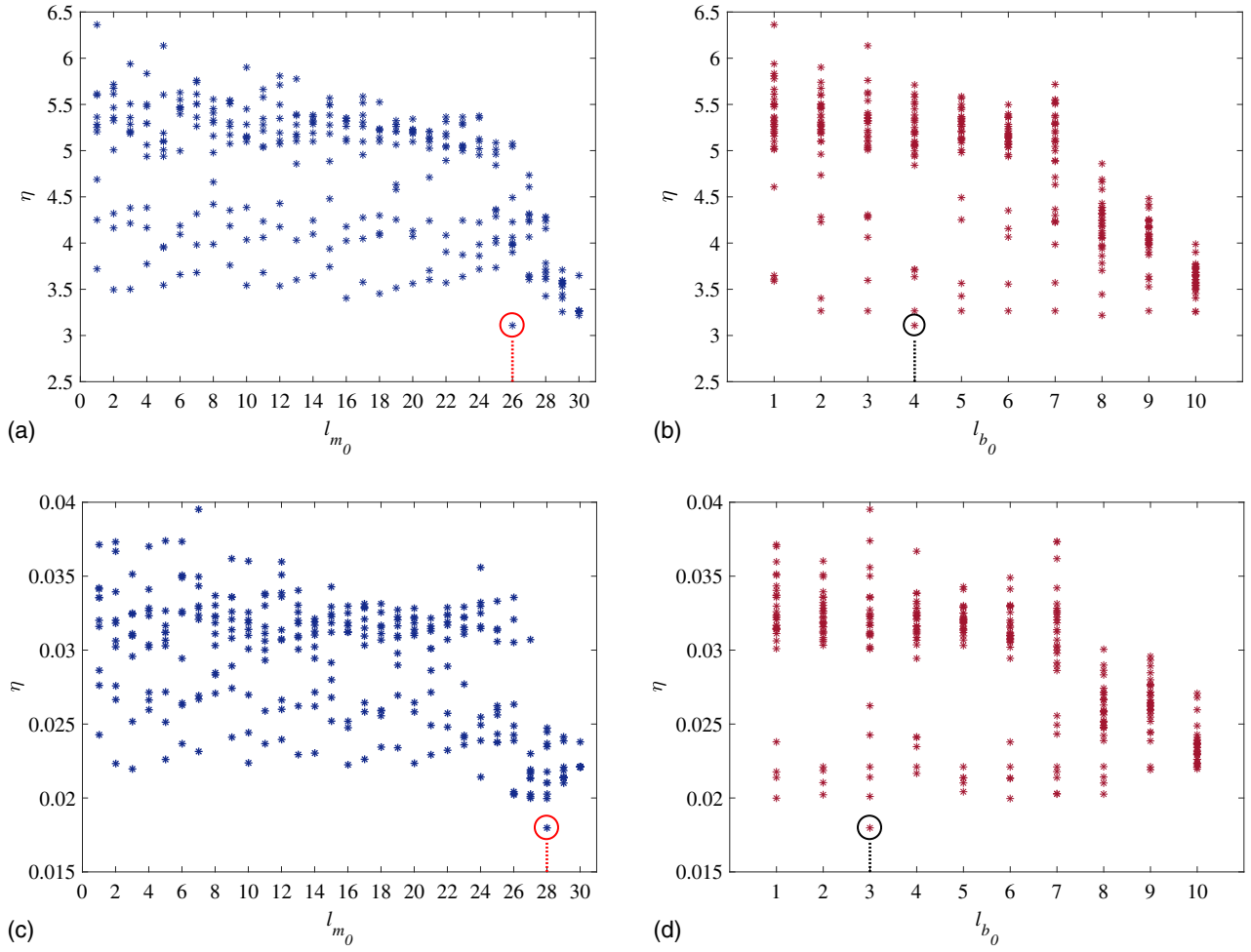

Fig. 5. Selection of the neuron sizes of the mapping, bottleneck, and demapping layers via the proposed automated hyperparameter optimization for the ANN (Algorithm 1): (a) l_m for ANN-MSD; (b) l_b for ANN-MSD; (c) l_m for ANN-SVD; and (d) l_b for ANN-SVD.

ANN-MSD method in conjunction with the EVT. The number and percentage of Type 2 errors in this method also were inconsiderable. After that, the proposed ANN-SVD method yielded better outputs than the nonparametric techniques. The clustering methods using the EVT provided small numbers and percentages of Type 1 error; however, the GMM outperformed the fuzzy *c*-means clustering in terms of Type 2 and total errors (Table 1). In the application of the CLT, the numbers and percentages of Type 2 errors of both the clustering methods were approximately in the same range and reasonable, whereas the Type 1 and total errors were considerable. In summary, the proposed semiparametric methods in conjunction with the EVT are the best choices for detecting damage under strong environmental variability. After them, one can utilize the GMM by considering the EVT.

Tianjin–Yonghe Bridge

The Tianjin–Yonghe Bridge is one of the earliest cable-stayed bridges with continuous prestressed box-girder constructed in China (Fig. 9). The total length of the Tianjin–Yonghe Bridge is 510 m, consisting of a main span of 260 m and two side spans of 25.15 and 99.85 m. The height of the two concrete towers of the bridge is 60.5 m, and they are connected by two transverse beams. After 19 years of operation, in 2005, serious cracks were found at the bottom of a girder segment over the midspan. Moreover, some cables close to the anchors were corroded severely. A major rehabilitation program was conducted to replace the damaged girder and all cables over the course of 2 years. In 2007, the bridge was monitored by a sophisticated SHM system organized by the Center of Structural Monitoring and Control (SMC) at the Harbin Institute of Technology in China. Nevertheless, new damage patterns were identified in the girders of the bridge during a routine inspection in August 2008.

From the report of the SMC group (Li et al. 2014), the vibration time-domain responses (acceleration time histories) on 12 days (i.e., January 1, January 17, February 3, March 19, March 30, April 19, May 5, May 18, May 31, June 7, June 16, and July 31) are available for use in damage detection. The vibration measurements include 24 sets of 1-h acceleration time histories with 360,000 data points/measurement (h) acquired from 14 single-axis accelerometers for 24 h. The sampling frequency and time interval of the acceleration responses were 100 Hz and 0.01 s, respectively. The initial analysis of the vibration responses indicated that the acceleration time histories of one of the accelerometers (Sensor 10) were meaningless and should be neglected. Moreover, the data for 3 days—May 31, June 7, and June 16—were excluded from the comparative study due to poor excitation conditions or lack of stability in the consecutive sets (Entezami et al. 2020a). In summary, the vibration responses from Sensors 1–9 and 11–14 for 9 days (i.e., January 1, January 17, February 3, March 19, March 30, April 19, May 5, May 18, and July 31) were considered in such a way that the vibration data of the first 8 days were associated with the normal

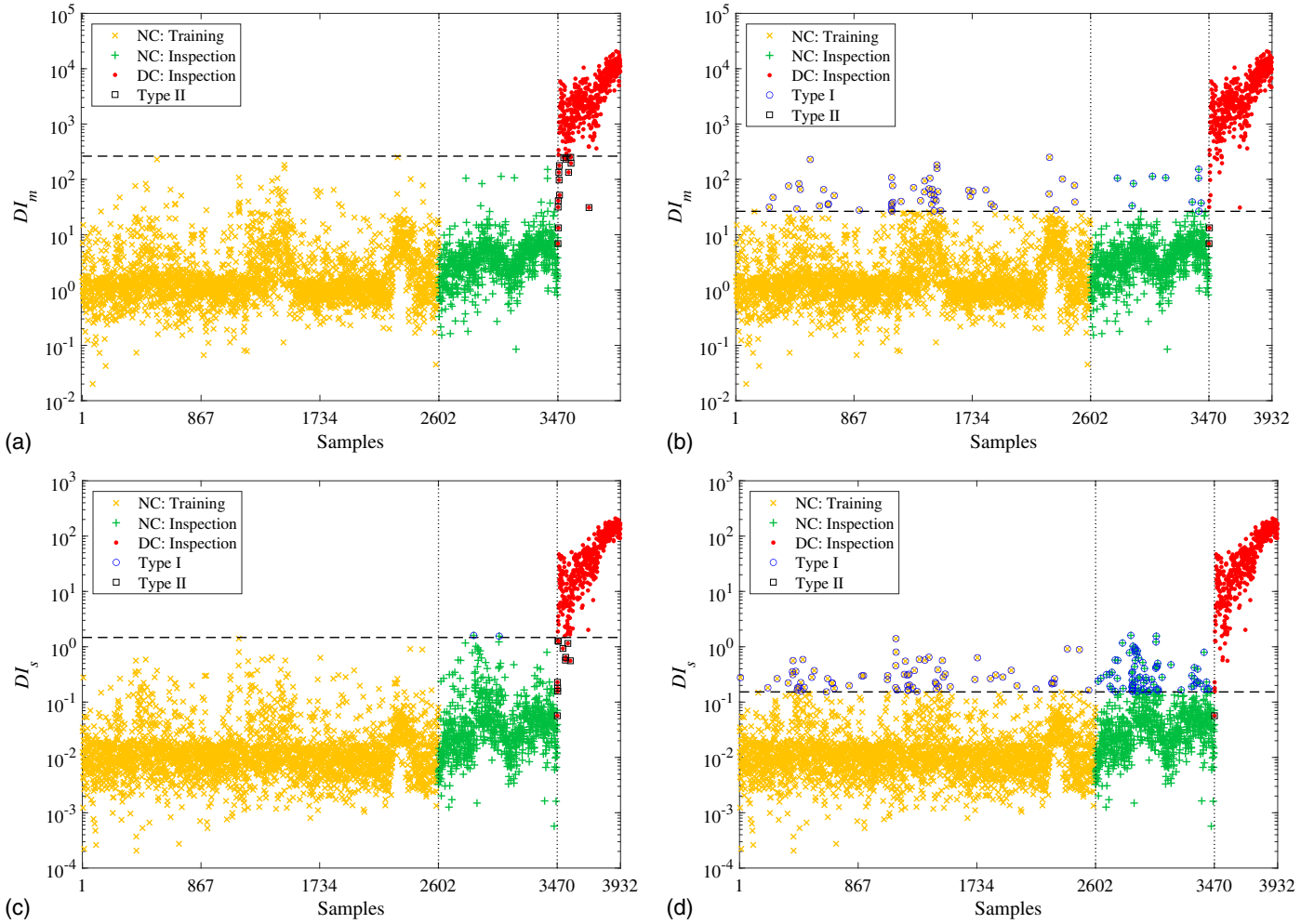

Fig. 6. Damage detection of the Z24 Bridge using the semiparametric methods: (a) ANN-MSD and EVT; (b) ANN-MSD and CLT; (c) ANN-SVD and EVT; and (d) ANN-SVD and CLT.

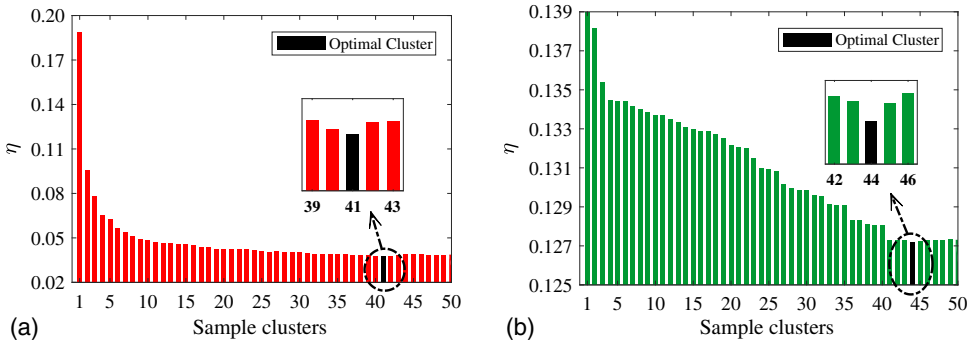

Fig. 7. Selection of the cluster numbers via the proposed automated method of hyperparameter optimization (Algorithm 2): (a) fuzzy *c*-means; and (b) GMM.

condition and those of the last day pertained to the damaged state of the bridge (Li et al. 2014).

In this case study, the main features of damage detection were extracted from the ARMA model. These features were the variances of the model residuals at all sensors and test measurements. The outputs of the ARMA modeling associated with the Tianjin–Yonghe Bridge are available in Entezami et al. (2020a). Accordingly, the variances of the ARMA residuals from 13 accelerometers and 24 test measurements of the first 8 days were calculated to produce a feature matrix of size 13×192 ($192 = 24 \times 8$) for the normal condition. The orders and coefficients of the ARMA models obtained from the first 8 days were used to extract the model residuals by considering the vibration responses of Day 9 and then calculating their variances. Hence, a feature matrix of the damaged state of the bridge the same size as the corresponding matrix related to the normal condition was obtained.

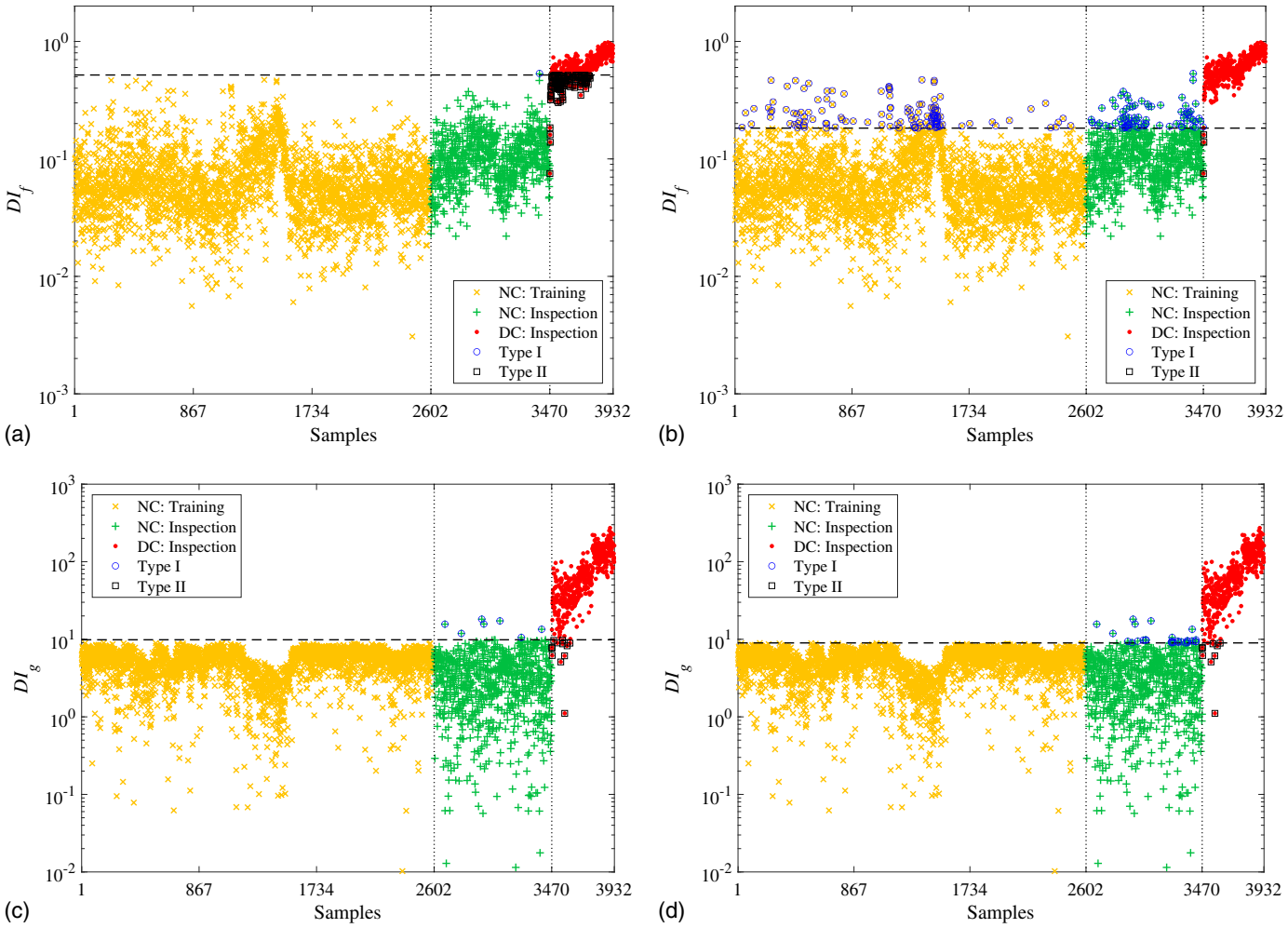

Fig. 8. Damage detection of the Z24 Bridge using the parametric methods: (a) fuzzy *c*-means clustering and EVT; (b) fuzzy *c*-means clustering and CLT; (c) GMM by EVT; and (d) GMM by CLT.

Table 1. Numbers and percentages of Type 1, Type 2, and total errors in detecting damage of Z24 Bridge using machine learning methods and two threshold estimation techniques

Method	Threshold	Type 1	Type 2	Total
MSD	EVT	0 (0)	462 (100)	462 (11.74)
	CLT	174 (5.01)	210 (45.45)	384 (9.76)
SVD	EVT	0 (0)	462 (100)	462 (11.74)
	CLT	138 (3.97)	277 (59.95)	415 (10.55)
ANN-MSD	EVT	0 (0)	7 (1.51)	7 (0.17)
	CLT	229 (6.59)	2 (0.43)	231 (5.87)
ANN-SVD	EVT	2 (0.057)	11 (2.38)	13 (0.33)
	CLT	145 (4.17)	1 (0.21)	146 (3.71)
Fuzzy *c*-means	EVT	1 (0.03)	129 (27.92)	130 (3.30)
	CLT	210 (6.05)	3 (0.65)	213 (5.41)
GMM	EVT	7 (0.20)	12 (2.59)	19 (0.48)
	CLT	28 (0.80)	9 (1.94)	37 (0.94)

Note: Values in parentheses are percentages (%).

Before detecting damage, it is important to predict the level of the EOV conditions (Fig. 10). For this process, all feature samples for the normal state (i.e., the feature matrix of size 13×192) and 1,000 sample repetitions were set as the inputs. The MAD values of the first 1,000 repetitions for the first stage varied in the range 0.080–0.086, which are relatively small amounts. In the second 1,000 repetitions, associated with the second stage, the MAD values were about 0.014–0.017, close to the corresponding values of the first stage. The ratio of the averages of the MAD amounts of the first to the second stages was 5.722. This relatively small amount and the distance between the MAD values demonstrate that the EOV conditions were not severe as those of the problem of the Z24 Bridge. Therefore, there was a weak level of EOV or low sensitivity of the ARMA residual variances to these conditions.

To detect damage via the machine learning methods, the training data comprised 75% of the residual variances of the normal condition; that is, $\mathbf{X} \in \mathbb{R}^{4\times144}$. The remaining 25% of the residual variances of the normal condition, along with all variances of the damaged state, were used to produce the test data $\mathbf{Z} \in \mathbb{R}^{4\times240}$. Using these matrices, Fig. 11 indicates the results of damage detection by the nonparametric methods through the EVT [Figs. 11(a and c)] and CLT-based [Figs. 11(b and d)] threshold estimation. All DI values of the training samples fell under the threshold lines, and a few points of the validation samples exceeded the thresholds. Moreover, the majority of the DI values of the damaged state in Samples 193–384 were over the threshold limits, indicating accurate damage detection, except for the two and three points (Type 2) associated with the MSD and SVD methods, respectively. In contrast, in Figs. 11(b and d), several DI values of the training and validation samples

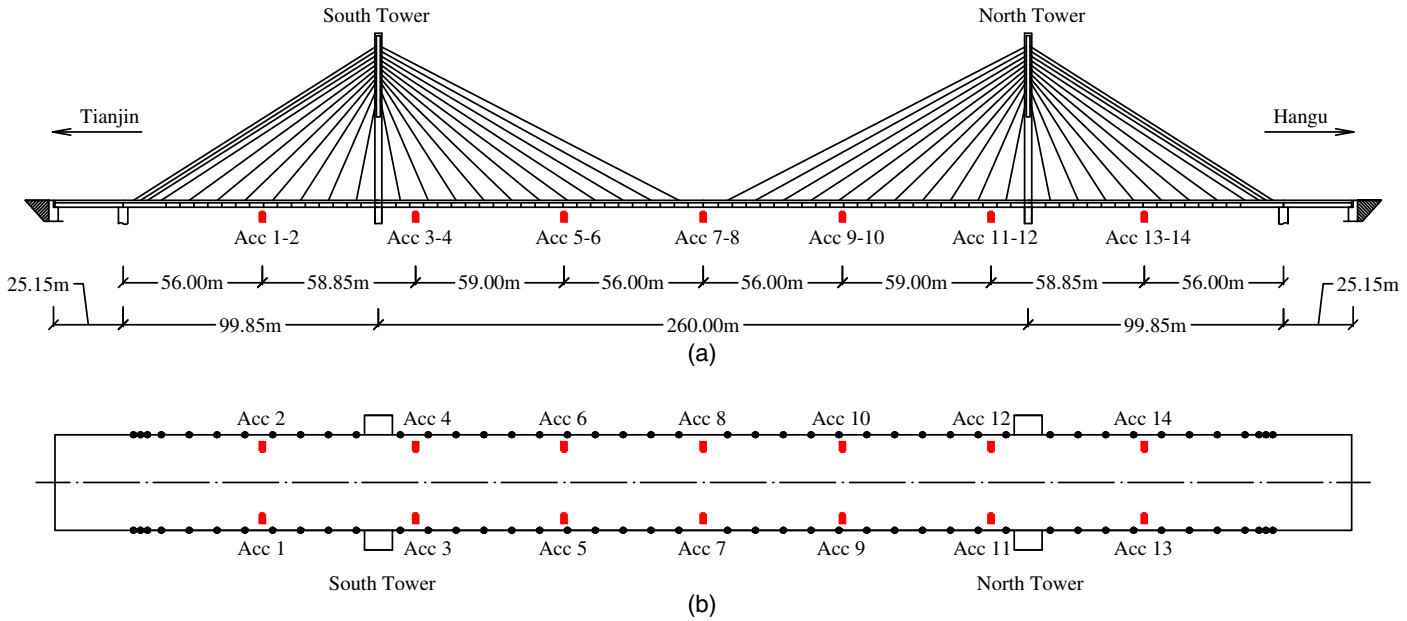

Fig. 9. Tianjin–Yonghe Bridge: (a) elevation view, main dimensions, and accelerometer labels; and (b) bridge plan and accelerometer locations.

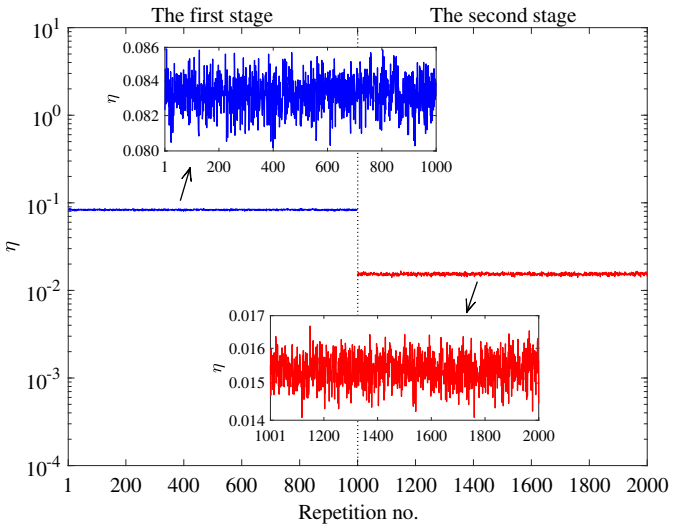

Fig. 10. Prediction the level of the EOV conditions in the Tianjin–Yonghe Bridge.

incorrectly exceeded the threshold lines (Type 1), and there was no Type 2 error.

Although the results of damage detection via the nonparametric methods and EVT were reasonable, potentially due to the weak EOV or low sensitivity of the extracted statistical features (i.e., the ARMA residual variances) to these conditions, it is appropriate to investigate the performance of the proposed semiparametric methods. Using the proposed automated hyperparameter optimization for the semiparametric methods, Fig. 12 illustrates the MAD values for training two autoassociative neural networks related to the ANN-MSD and ANN-SVD methods, in which the sample neurons l_{m_0} and l_{b_0} were set to 30 and 10, respectively. The optimal neurons of the mapping and demapping layers (l_m) for the ANN-MSD and ANN-SVD methods were 15 and 13, respectively [Figs. 12(a and c)]. Moreover, the optimal neurons of the bottleneck layer (l_b) were 2 for both methods [Figs. 12(b and d)].

Utilizing these neurons, without any overfitting problem, and training two autoassociative neural networks, the residual matrices $\mathbf{E_x}$ and $\mathbf{E_z}$ were extracted and applied to the MSD and SVD equations. The results of damage detection using the proposed semiparametric methods are shown in Fig. 13, where the horizontal dashed lines refer to the thresholds estimated by the EVT [Figs. 13(a and c)] and the 95% confidence interval based on the CLT [Figs. 13(b and d)] at a 5% significance level. In Figs. 13(a and c), there are no Type 1 and Type 2 errors. This means that the ANN in conjunction with the EVT highly assisted the nonparametric techniques to decrease the variations in the DI values of the normal condition and deal with any outlier masking problem, leading to better results than direct using the MSD and SVD methods. In contrast, in Figs. 13(b and d), despite proper damage detectability and the removal of EOV, the semiparametric methods suffered from Type 1 errors when the threshold estimation was based on the CLT and the standard confidence interval.

For damage detection via the clustering techniques, Fig. 14 displays the cluster numbers obtained from the proposed automated approach. On this basis, the optimal cluster numbers c and r associated with the fuzzy c-means and GMM were 16, and 11, respectively. Using these numbers and implementing the clustering algorithms, the results of damage detection are shown in Fig. 15, where the horizontal dashed lines are the thresholds obtained from the EVT [Figs. 15(a and c)] and the 95% confidence interval based on the CLT [Figs. 15(b and d)]. In Figs. 15(a and c), no DI values of the training samples exceeded the threshold lines, and only two DI quantities of the validation samples associated with the GMM were over the thresholds. In contrast, the number of Type 1 errors increases in both clustering methods by considering the CLT-based threshold estimation. Hence, this conclusion confirms the limitation of this technique for estimating a reliable alarming threshold even under weak EOV conditions. Regarding the Type 2 error, expect for the fuzzy c-means [Figs. 15(a and b)], it is seen that all DI values of Samples 193–384 related to the GMM [Figs. 15(c and d)] were above the thresholds. This implies the superiority of the GMM over the fuzzy c-means clustering in the problem of damage detection.

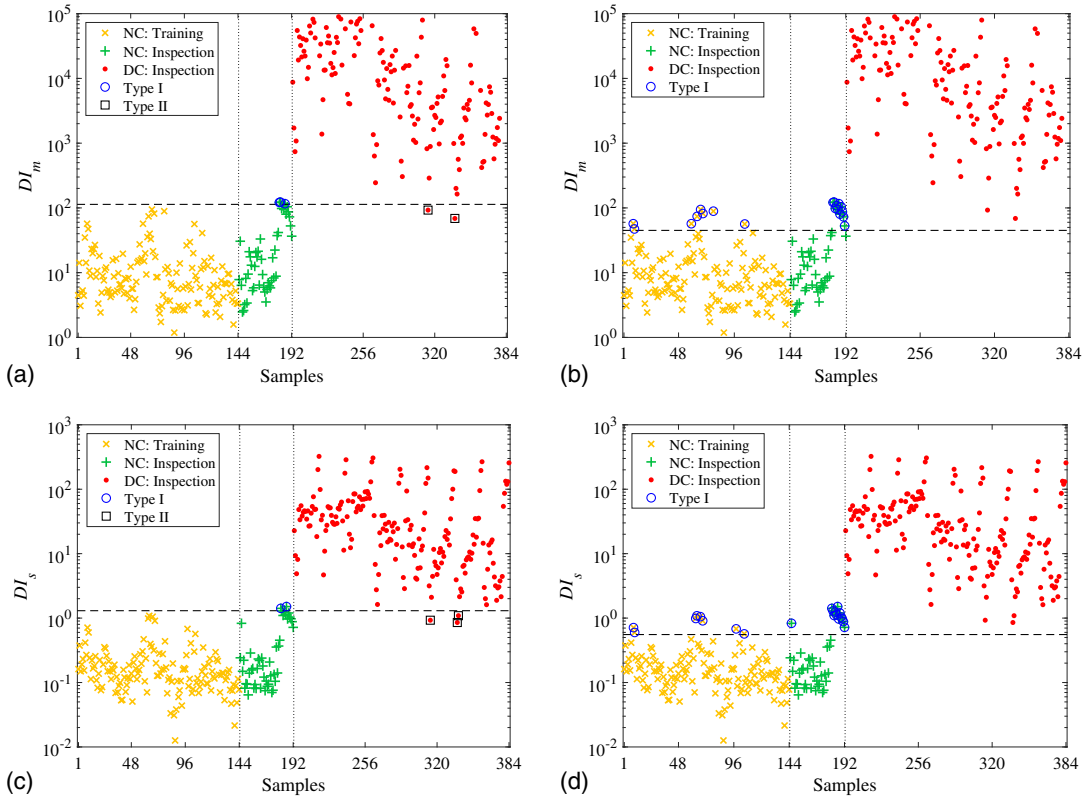

Fig. 11. Damage detection of the Tianjin–Yonghe Bridge using the nonparametric methods: (a) MSD and EVT; (b) MSD and CLT; (c) SVD and EVT; and (d) SVD and CLT.

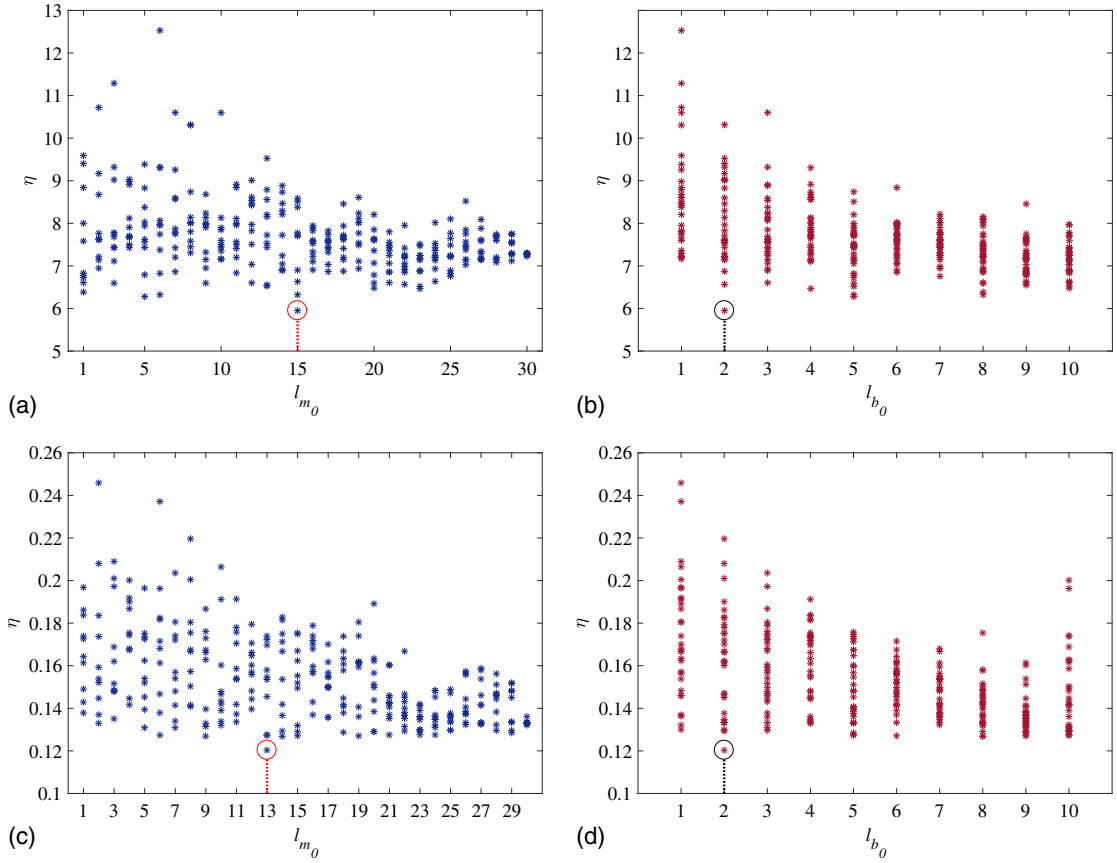

Fig. 12. Selection of the neuron sizes of the mapping, bottleneck, and demapping layers via the proposed automated hyperparameter optimization for the ANN (Algorithm 1): (a) l_m for ANN-MSD; (b) l_b for ANN-MSD; (c) l_m for ANN-SVD; and (d) l_b for ANN-SVD.

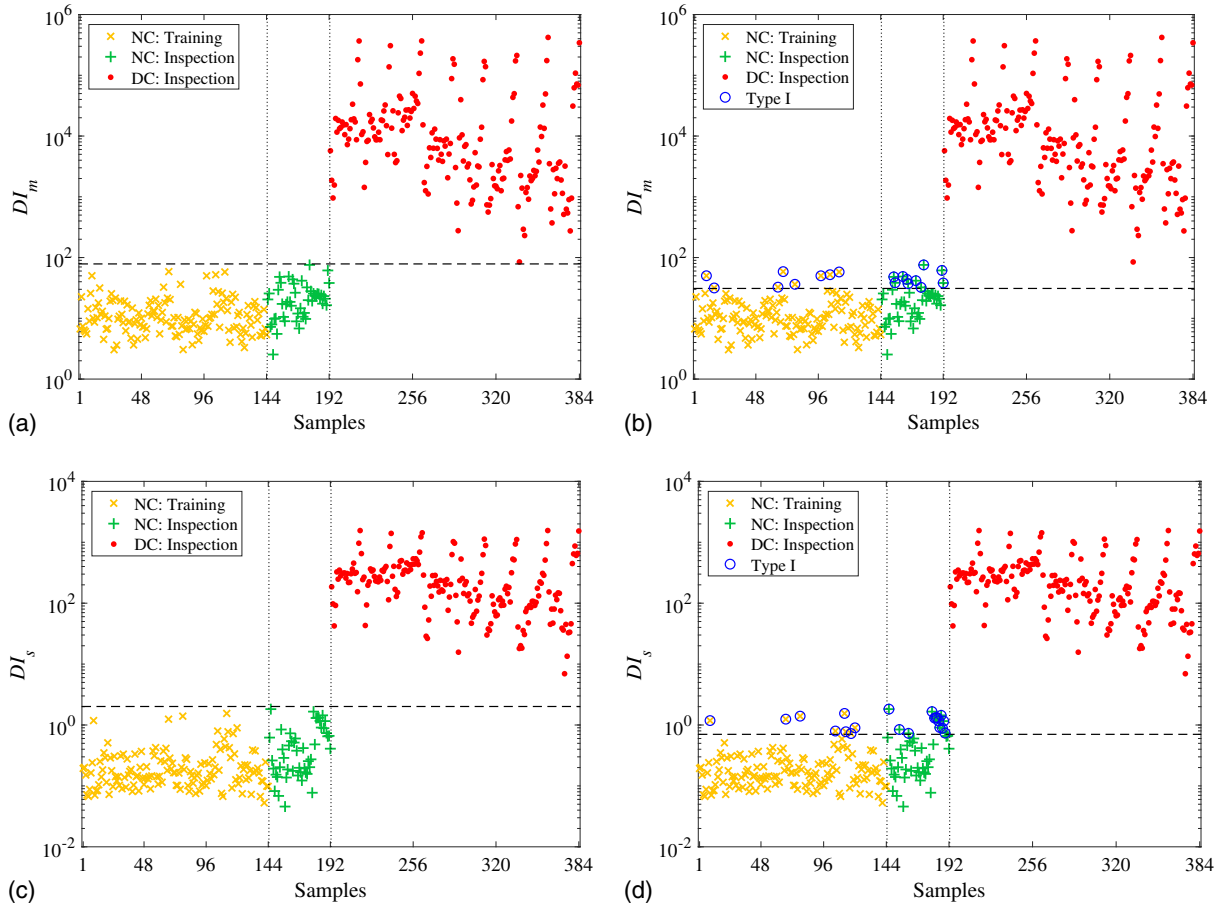

Fig. 13. Damage detection of the Tianjin–Yonghe Bridge using the semiparametric methods: (a) ANN-MSD and EVT; (b) ANN-MSD and CLT; (c) ANN-SVD and EVT; and (d) ANN-SVD and CLT.

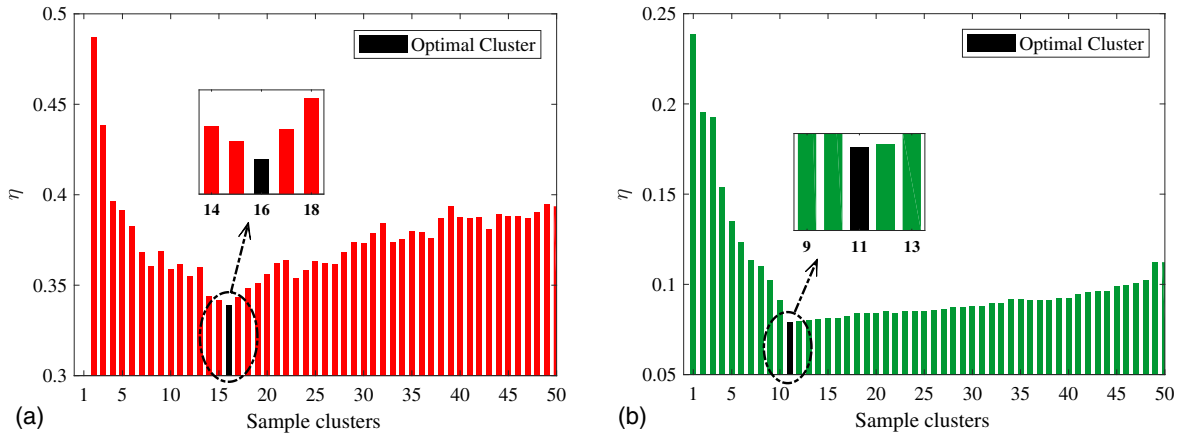

Fig. 14. Selection of the cluster numbers via the proposed automated method of hyperparameter optimization (Algorithm 2): (a) fuzzy c-means; and (b) GMM.

For further evaluation, Table 2 lists the numbers and percentages of Type 1, Type 2, and total errors in detecting damage via the nonparametric, semiparametric, and parametric methods in conjunction with the EVT and CLT. Unlike in the previous case study, the error rates in Table 2 for the MSD and SVD methods are small. The excellent performance was related to the semiparametric methods based on the EVT, for which there were no errors. Of the clustering techniques, the GMM outperformed the fuzzy c-means clustering using the EVT. With the use of the CLT, both the clustering methods yielded approximately the same results, which were worse than the corresponding results obtained using the EVT. In summary, the proposed semiparametric methods with the aid of the EVT are the best choices for damage detection. The GMM is the next best technique. This important conclusion is related to the possibility of

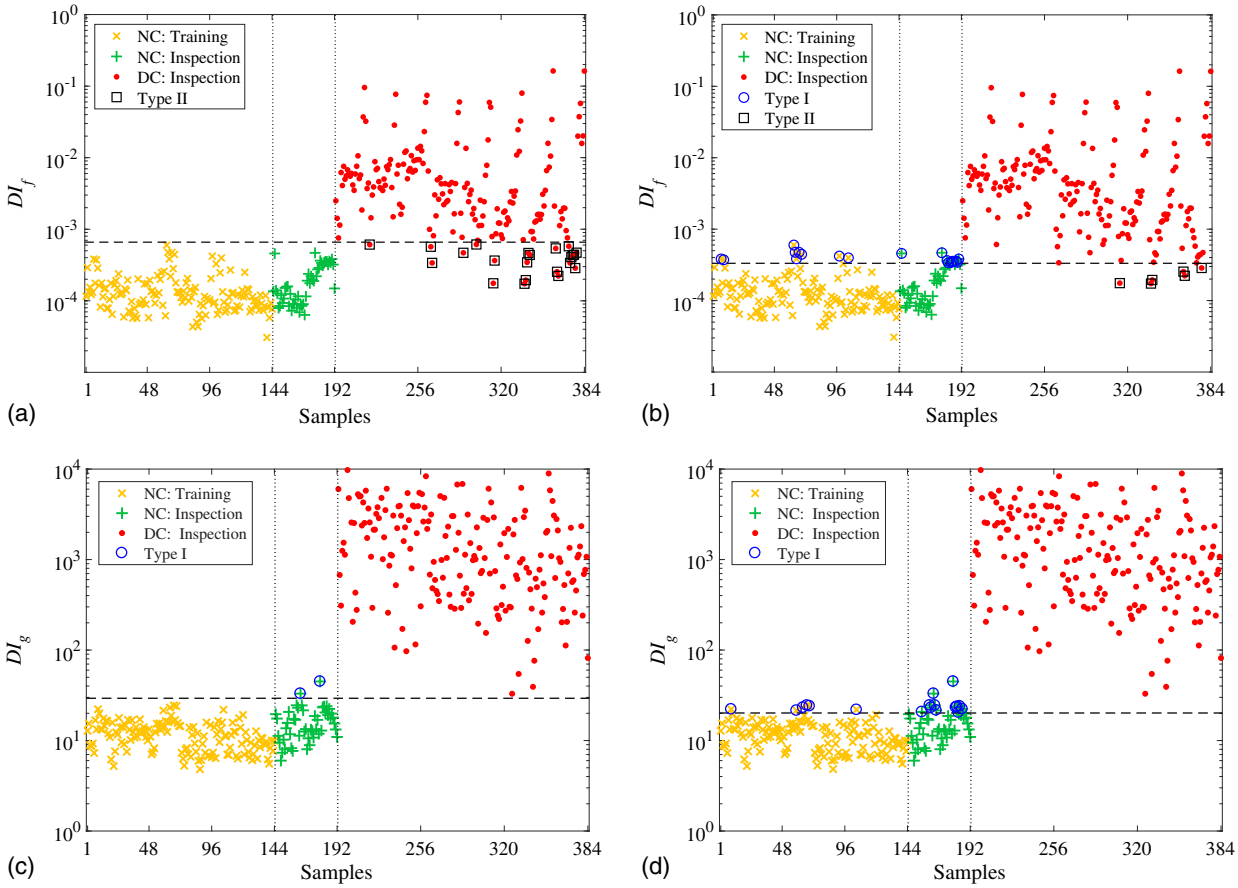

Fig. 15. Damage detection of the Tianjin–Yonghe Bridge using the parametric methods: (a) Fuzzy c-means clustering and EVT; (b) Fuzzy c-means clustering and CLT; (c) GMM by EVT; and (d) GMM by CLT.

Table 2. Numbers and percentages of Type 1, Type 2, and total errors in detecting damage using machine learning methods and two threshold estimation techniques

Method	Threshold	Type 1	Type 2	Total
MSD	EVT	3 (1.56)	2 (1.04)	5 (1.30)
	CLT	19 (9.89)	0 (0)	19 (4.94)
SVD	EVT	0 (0)	3 (1.56)	3 (0.78)
	CLT	14 (7.29)	0 (0)	14 (3.64)
ANN-MSD	EVT	0 (0)	0 (0)	0 (0)
	CLT	18 (9.37)	0 (0)	18 (4.68)
ANN-SVD	EVT	0 (0)	0 (0)	0 (0)
	CLT	21 (10.93)	0 (0)	21 (5.46)
Fuzzy c-means	EVT	0 (0)	22 (11.45)	22 (5.72)
	CLT	19 (9.89)	6 (3.12)	25 (6.51)
GMM	EVT	2 (1.04)	0 (0)	2 (0.52)
	CLT	20 (10.41)	0 (0)	20 (5.21)

Note: Values in parentheses are percentages (%).

using the nonparametric methods based on the EVT in the weak EOV conditions. This confirms the importance of predicting the level of these conditions before SHM.

Conclusions

Due to the importance of SHM for ensuring the safety and integrity of civil structures, this article conducted a comprehensive comparative study of various machine learning methods under unsupervised learning to detect damage under different EOV conditions, and proposed new approaches to the problem of hyperparameter optimization. The machine learning methods included the nonparametric MSD and SVD, the semiparametric ANN-MSD and ANN-SVD, and the parametric fuzzy c-means and GMM clustering algorithms. In the comparative study, two alarming thresholds based on the EVT using the generalized extreme value distribution and the CLT with the standard confidence interval were applied to make decisions about the occurrence of damage. The new approaches were automated hyperparameter optimization algorithms based on the MAD criterion for choosing the neuron sizes of the hidden layers of an autoassociative neural network and the cluster numbers of the clustering techniques by the focus on dealing with the outlier masking problem caused by the EOV conditions. An automated output-only method using the MAD criterion was also proposed to predict the level of the EOV. Finally, the performance of the machine learning methods, the proposed approaches, and the threshold estimation techniques was evaluated and compared in terms of the dynamic and statistical features of two full-scale bridge structures.

The main conclusions of this article can be summarized as follows:

1. In all comparisons, the threshold estimation via the CLT and standard confidence interval was not successful in providing reliable results. This may be due to the non-Gaussian distribution of the DI values used for the threshold estimation. Hence, this technique is not recommended for use in threshold estimation. In contrast, the EVT yielded the best performance in terms of the smallest rates of Type 1, Type 2, and total errors.

Therefore, one can apply it to determine a reliable alarming threshold.

2. The nonparametric MSD and SVD methods failed to provide accurate results of damage detection under the strong EOV related to the Z24 Bridge. However, they were successful under the weak EOV conditions regarding the SHM problem of the Tianjin–Yonghe Bridge. This conclusion confirms the importance of predicting the level of the EOV conditions and analyzing any kind of features extracted from measured vibration data.
3. In both the strong and weak EOV conditions, the best performance was that of the proposed semiparametric ANN-MSD and ANN-SVD methods in conjunction with the EVT. This conclusion also proves the effectiveness of the proposed hyperparameter optimization regarding the neuron size selection and the application of the autoassociative neural network to properly addressing the outlier masking problem.
4. Among the parametric methods, the GMM outperformed the fuzzy c-means clustering in both the strong and weak EOV conditions. However, the semiparametric methods provided better results than this technique.

For further studies, it is recommended to study advanced hyperparameter optimization algorithms such as gradient-based, Bayesian, and multifidelity optimizations, as well as metaheuristic approaches to addressing the problem of machine learning models with large hyperparameters. This important subject can be merged with advanced machine learning algorithms based on the concepts of semisupervised learning, deep learning, transfer learning, active learning, ensemble learning, and kernel learning.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

References

Agdas, D., J. A. Rice, J. R. Martinez, and I. R. Lasa. 2016. "Comparison of visual inspection and structural-health monitoring as bridge condition assessment methods." *J. Perform. Constr. Facil.* 30 (3): 04015049. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000802.

Aggarwal, C. C., and C. K. Reddy. 2016. *Data clustering: Algorithms and applications*. Boca Raton, FL: CRC Press.

Avci, O., and O. Abdeljaber. 2016. "Self-organizing maps for structural damage detection: A novel unsupervised vibration-based algorithm." *J. Perform. Constr. Facil.* 30 (3): 04015043. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000801.

Bagchi, A., J. Humar, H. Xu, and A. S. Noman. 2010. "Model-based damage identification in a continuous bridge using vibration data." *J. Perform. Constr. Facil.* 24 (2): 148–158. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000071.

Chang, C.-M., T.-K. Lin, and C.-W. Chang. 2018. "Applications of neural network models for structural health monitoring based on derived modal properties." *Measurement* 129 (7): 457–470. https://doi.org/10.1016/j.measurement.2018.07.051.

Daneshvar, M. H., A. Gharighoran, S. A. Zareei, and A. Karamodin. 2021. "Early damage detection under massive data via innovative hybrid methods: application to a large-scale cable-stayed bridge." *Struct. Infrastruct. Eng.* 17 (7): 902–920. https://doi.org/10.1080/15732479.2020.1777572.

Deraemaeker, A., E. Reynders, G. De Roeck, and J. Kullaa. 2008. "Vibration-based structural health monitoring using output-only measurements under changing environment." *Mech. Syst. Sig. Process.* 22 (1): 34–56. https://doi.org/10.1016/j.ymssp.2007.07.004.

Dodge, Y. 2008. *The concise encyclopedia of statistics*. New York: Springer.

Entezami, A., H. Sarmadi, B. Behkamal, and S. Mariani. 2020a. "Big data analytics and structural health monitoring: A statistical pattern recognition-based approach." *Sensors* 20 (8): 2328. https://doi.org/10.3390/s20082328.

Entezami, A., H. Sarmadi, M. Salar, C. De Michele, and A. Nadir Arslan. 2021. "A novel data-driven method for structural health monitoring under ambient vibration and high-dimensional features by robust multidimensional scaling." *Struct. Health Monit.* 2021 (1): 1475921720973953. https://doi.org/10.1177/1475921720973953.

Entezami, A., H. Shariatmadar, and S. Mariani. 2020b. "Early damage assessment in large-scale structures by innovative statistical pattern recognition methods based on time series modeling and novelty detection." *Adv. Eng. Software* 150 (Dec): 102923. https://doi.org/10.1016/j.advengsoft.2020.102923.

Farrar, C. R., and K. Worden. 2013. *Structural health monitoring: A machine learning perspective*. West Sussex, UK: Wiley.

Feurer, M., and F. Hutter. 2019. "Hyperparameter optimization." In *Automated machine learning*, 3–33. Berlin: Springer.

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

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

Flah, M., I. Nunez, W. Ben Chaabene, and M. L. Nehdi. 2020. "Machine learning algorithms in civil structural health monitoring: A systematic review." *Arch. Comput. Methods Eng.* 28 (4): 2621–2643. https://doi.org/10.1007/s11831-020-09471-9.

Gharibnezhad, F., L. E. Mujica, and J. Rodellar. 2015. "Applying robust variant of Principal Component Analysis as a damage detector in the presence of outliers." *Mech. Syst. Sig. Process.* 50-51 (Jan): 467–479. https://doi.org/10.1016/j.ymssp.2014.05.032.

Jeong, E., J. Seo, and J. Wacker. 2020. "Literature review and technical survey on bridge inspection using unmanned aerial vehicles." *J. Perform. Constr. Facil.* 34 (6): 04020113. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001519.

Kramer, M. A. 1992. "Autoassociative neural networks." *Comput. Chem. Eng.* 16 (4): 313–328. https://doi.org/10.1016/0098-1354(92)80051-A.

Leys, C., C. Ley, O. Klein, P. Bernard, and L. Licata. 2013. "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median." *J. Exp. Social Psychol.* 49 (4): 764–766. https://doi.org/10.1016/j.jesp.2013.03.013.

Li, S., H. Li, Y. Liu, C. Lan, W. Zhou, and J. Ou. 2014. "SMC structural health monitoring benchmark problem using monitored data from an actual cable-stayed bridge." *Struct. Contr. Health Monit.* 21 (2): 156–172. https://doi.org/10.1002/stc.1559.

Liu, W., X. Wu, L. Zhang, Y. Wang, and J. Teng. 2020. "Structural health-monitoring and assessment in tunnels: hybrid simulation approach." *J. Perform. Constr. Facil.* 34 (4): 04020045. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001445.

McLachlan, G. J., and T. Krishnan. 2007. *The EM algorithm and extensions*. Hoboken, NJ: Wiley.

Reynders, E., and G. De Roeck. 2009. "Encyclopedia of structural health monitoring." In *Continuous vibration monitoring and progressive damage testing on the Z24 bridge*. Hoboken, NJ: Wiley.

Ruotolo, R., and C. Surace. 1999. "Using SVD to detect damage in structures with different operational conditions." *J. Sound Vib.* 226 (3): 425–439. https://doi.org/10.1006/jsvi.1999.2305.

Sarmadi, H., and A. Entezami. 2021. "Application of supervised learning to validation of damage detection." *Arch. Appl. Mech.* 91 (1): 393–410. https://doi.org/10.1007/s00419-020-01779-z.

Sarmadi, H., A. Entezami, B. Saeedi Razavi, and K.-V. Yuen. 2021a. "Ensemble learning-based structural health monitoring by Mahalanobis distance metrics." *Struct. Control Health Monit.* 28 (2): e2663. https://doi.org/10.1002/stc.2663.

Sarmadi, H., A. Entezami, M. Salar, and C. De Michele. 2021b. "Bridge health monitoring in environmental variability by new clustering and

threshold estimation methods." *J. Civ. Struct. Health Monit.* 11 (3): 629–644. https://doi.org/10.1007/s13349-021-00472-1.

Sarmadi, H., and A. Karamodin. 2020. "A novel anomaly detection method based on adaptive Mahalanobis-squared distance and one-class kNN rule for structural health monitoring under environmental effects." *Mech. Syst. Sig. Process.* 140 (4): 106495. https://doi.org/10.1016/j.ymssp.2019.106495.

Sarmadi, H., and K.-V. Yuen. 2021. "Early damage detection by an innovative unsupervised learning method based on kernel null space and peak-over-threshold." *Copmut. Aided Civil Infrastruct. Eng*. 2021 (May): 5. https://doi.org/10.1111/mice.12635.

Seo, J., J. W. Hu, and J. Lee. 2016. "Summary review of structural health monitoring applications for highway bridges." *J. Perform. Constr. Facil.* 30 (4): 04015072. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000824.

Sun, L., Z. Shang, Y. Xia, S. Bhowmick, and S. Nagarajaiah. 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.

Weinstein, J. C., M. Sanayei, and B. R. Brenner. 2018. "Bridge damage identification using artificial neural networks." *J. Bridge Eng.* 23 (11): 04018084. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001302.

Yang, L., and A. Shami. 2020. "On hyperparameter optimization of machine learning algorithms: Theory and practice." *Neurocomputing* 415 (2): 295–316. https://doi.org/10.1016/j.neucom.2020.07.061.

Yang, Y., C. Dorn, T. Mancini, Z. Talken, G. Kenyon, C. Farrar, and D. Mascareñas. 2017. "Blind identification of full-field vibration modes from video measurements with phase-based video motion magnification." *Mech. Syst. Sig. Process.* 85 (3): 567–590. https://doi.org/10.1016/j.ymssp.2016.08.041.

Yekrangnia, M., and A. A. Mobarake. 2016. "Restoration of historical Al-Askari shrine. I: Field observations, damage detection, and material properties." *J. Perform. Constr. Facil.* 30 (3): 04015030. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000761.

Yu, L., and J. H. Zhu. 2017. "Structural damage prognosis on truss bridges with end connector bolts." *J. Eng. Mech.* 143 (3): B4016002. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001052.