IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 12 June 2023, accepted 26 June 2023, date of publication 3 July 2023, date of current version 7 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3291674

RESEARCH ARTICLE

Semi-Supervised Detection of Structural Damage Using Variational Autoencoder and a One-Class Support Vector Machine

ANDREA POLLASTRO[1], GIUSIANA TESTA[2], ANTONIO BILOTTA[2], AND ROBERTO PREVETE[1]

[1]Department of Electrical Engineering and Information Technology, University of Naples Federico II, 80125 Naples, Italy
[2]Department of Structures for Engineering and Architecture, University of Naples Federico II, 80125 Naples, Italy

Corresponding author: Andrea Pollastro (andrea.pollastro@unina.it)

This work is supported by PRIN research project "BRIO – BIAS, RISK, OPACITY in AI: design, verification and development of Trustworthy AI.", Project no. 2020SSKZ7R. Furthermore, we acknowledge financial support from the Piano Nazionale di Ripresa e Resilienza (PNRR), Ministero dell'Universita e della Ricerca (MUR) project PE0000013-FAIR and ReLUIS Ponti Project funded by Consiglio Superiore dei Lavori Pubblici (CSLP) of the Italian Infrastructure Ministry.

ABSTRACT In recent years, Artificial Neural Networks (ANNs) have been introduced in Structural Health Monitoring (SHM) systems. A semi-supervised method with a data-driven approach allows the ANN training on data acquired from an undamaged structural condition to detect structural damages. In standard approaches, after the training stage, a decision rule is manually defined to detect anomalous data. However, this process could be made automatic using machine learning methods. This paper proposes a semi-supervised method with a data-driven approach to detect structural anomalies. The methodology consists of: 1) a Variational Autoencoder (VAE) to approximate undamaged data distribution and 2) a One-Class Support Vector Machine (OC-SVM) to discriminate different health conditions using damage-sensitive features extracted from VAE's signal reconstruction. The method is applied to a scale steel structure that was tested in nine damage scenarios by IASC-ASCE Structural Health Monitoring Task Group.

INDEX TERMS Semi-supervised damage detection, structural health monitoring, variational autoencoder, one-class support vector machines, machine learning.

I. INTRODUCTION

Anomaly detection is a key research problem within many diverse research areas and application domains (see, for example, [1], [2], [3]). *Anomalies* (also said *abnormalities*, *deviants*, or *outliers*) can be viewed as data instances which move away, are dissimilar, from the large part of collected data. Errors in the data can be the cause of anomalies, but sometimes they can be indicative of a new, previously unknown, underlying process [4]. Anomaly detection tasks have been tackled by several Machine Learning (ML), and in particular Deep Learning (DL), techniques [5], [6], [7]. However, a substantial part of anomaly detection approaches is based on Autoencoder (AE) architectures [4], [8], [9], [10], [11], [12], [13]. AEs correspond to neural networks composed of at least one hidden layer and logically divided into two components, an *encoder* and a *decoder*. From a functional point of view, an AE can be seen as the composition of two functions E and D: E is an encoding function (the encoder) which maps the input space onto a feature space (or latent encoding space), D is a decoding function (the decoder) which inversely maps the feature space on the input space. A meaningful aspect is that by AEs, one can obtain data representations in terms of fixed latent encodings $\vec{h}$. In a nutshell, in anomaly detection tasks AEs are trained to minimize reconstruction error only on normal data instances, thus involving high reconstruction error on anomalous data. Then, the reconstruction error is considered as an anomaly score to classify the input data as anomalous or not, using a user-defined decision rule [14]. AEs' architectures have been

The associate editor coordinating the review of this manuscript and approving it for publication was Kathiravan Srinivasan.

presented with several variations such as Denoising Autoencoders (DAE), [15] which were meant to remove additional noise from input data, Sparse Autoencoders (SAE) [16], where a sparsity constraint is introduced on the hidden layer in order to emphasize meaningful features, and Variational Autoencoders (VAE) [17], that are generative models where the latent space is composed by a mixture of distributions instead of a fixed vector.

In recent decades, the attention to procedures for anomaly detection due to damage phenomena in civil constructions and infrastructures is more and more growing. Indeed, (i) safety standards for new constructions have increased - and therefore existing constructions could not comply with these standards for little degradation phenomena (ii) both new and existing structures are becoming increasingly smart with the use of several embedded sensors providing real-time information. For this reason, the research aimed at finding procedures that allow the set up of a Structural Health Monitoring (SHM) system for structures and infrastructures, i.e., for both buildings and bridges, are very numerous. Bridges are strategic structures for which important and expensive management and maintenance activities are foreseen because they are structural types particularly subject to environmental phenomena and variations in use conditions (loading-unloading cycles, temperature, etc.). Moreover, they do not have reserves of resistance capacity, which are characteristic of other structural types such as, for example, buildings. On the one hand, a proper model of the physics behavior of this type of structures in operational condition is not easy. This stimulates the use of automatic monitoring systems that can continuously and rapidly detect anomalous conditions due to damage, to ensure a quick response from the infrastructure manager. On the other hand, it is necessary to consider that (i) the high variability of the boundary conditions in which the bridge structure functions can alter the estimate of the anomaly (e.g., variable vibrations induced by wind actions, highly variable traffic load during the functioning of the structure, highly non-linear mechanical behavior of the materials that constitute the bridge) (ii) any algorithm implemented for a structural monitoring system hardly detect damage conditions if trained on an extensive database of measurements performed mainly in the operating conditions of the structure, namely in the absence of structural damage. This second aspect is crucial because the difficulties of measuring damage conditions are due to the intrinsic assumption made in the structural design approach, which expects the use of high safety factors to ensure that the operational conditions are well far from the structural limit condition. Therefore it is evident that investigating the use of damage detection algorithms that accurately provide warnings for structural monitoring is particularly challenging and interesting, regardless the subsequent necessity of damage quantification and structural prognostics. The monitoring strategies are mainly characterized by (i) types of monitoring (static or dynamic), (ii) analysis methodologies (i.e. input-output, with known forces, or output-only, with unknown forces) and (iii) analysis approach (i.e. data-driven or model-based, depending on whether the creation of a model to support the method is required). Static monitoring techniques usually consist of discrete more than continuous detection of gradual and slow variations of some parameters in rather long periods. By contrast, dynamic monitoring methodologies - which can use different techniques for identifying dynamic parameters, in the frequency domain [18] (e.g. peak picking, frequency domain decomposition, enhanced frequency domain decomposition) and in the time domain [19] (e.g. auto-regressive moving average models) - generally need to use a large amount of data. The records of accelerations, speeds and displacements can be post-processed through techniques operating in time or frequency domain, which affects the damage-sensitive feature. In the frequency domain, the features can be curvature, strain energy, flexibility and interpolation error [20], [21] while, in the time domain, the feature is generally an error parameter [22].

In this work, we propose a semi-supervised data-driven DL-based framework to detect damages in an SHM system. Our proposal consists in using a VAE, trained on undamaged raw data, to represent input data through *damage-sensitive* features (typically involved in structural damage detection [23], [24], [25]) and a One-Class Support Vector Machines (OC-SVM) [26] to classify data as undamaged or not, thus avoiding any user-defined decision rule. Damage-sensitive features are extracted by input data and their reconstruction computed through the VAE. Differently from other works based on standard AEs, our proposal leverages on the probabilistic aspects of a VAEs for the extraction of damage-sensitive features from input raw data, which implies the capturing of more data variability in the latent encoding space than a standard AE, avoiding in this way several weaknesses that may be found by using AEs for anomaly detection instead [14]. Moreover, since the probabilistic encoder of a VAE approximates the generative distribution of input data through their latent representation (differently from an AEs, where a deterministic mapping from the input to the latent representation is learnt [14]), we expect that learning the distribution of undamaged data lets the encoder to model damaged data with different distributions, thus improving the robustness of the damage detection system. Finally, to the best of our knowledge, among various anomaly diagnosis studies in SHM based on machine learning methods, this paper aims to propose for the first time an analysis of the VAE latent representations in modeling damaged/undamaged data distribution and its impact on the damage detection through KL divergence analysis on the various damage cases.

This paper is organized as follows. Section II briefly reviews the related literature; Section III describes the proposed architecture; Section IV introduces the experimental assessment together with the discussion about the results, while in Section V an analysis on the VAE's functioning

is provided. The concluding Section VIII is left to final remarks.

II. RELATED WORKS

During the last years, due to the great success achieved in solving several kinds of problems and due to the increasing accessibility to computing hardware, the interest in using DL-based approach in processing massive data coming from SHM systems is raising, thus moving researchers to design SHM damage detection methodologies towards autonomous data-driven systems. One of the main advantages of introducing DL methods in SHM systems consists in automating the feature extraction process from raw input data through learnable non-linear transformations modeled as layers of a Deep Neural Network (DNN), thus eliminating the need for human-designed features, the requirement for specific feature knowledge and resulting in a DL-based SHM system that is end-to-end. [27]. The use of DNNs has introduced the possibility to process large datasets acquired from different types of sensors in data-driven SHM systems [28], [29].

Yan et al. in [30] presented a multiscale cascading deep belief network named MCDBN for automatic fault identification of rotating machinery. The same authors in [31] proposed a novel hybrid deep learning model for multistep forecasting of diurnal wind speed called ISSD-LSTM-GOASVM. In [32], Xu et al. provided a summary of the state-of-the-art progress of AI applications in civil engineering for the entire life cycle of civil infrastructures. Li et al. in [33] conducted a comparison between the performance of a Convolutional Neural Network (CNN) and other methods, such as Support Vector Machine, Random Forest, k-Nearest Neighbor, and Decision Trees for damage detection in an experimental cable bridge model. The results demonstrated that the accuracy score was improved by at least 15 % when using a CNN. In [34], Li et al. presented an approach that integrates the electromechanical admittance (EMA) technique with CNNs to quantify structural damage severity under varied temperatures. Ai et al. in [35] proposed a novel approach based on CNNs integrated with EMA to identify compressive stress and load-induced damages of concrete cubic structures subjected to loading. The same authors, in [36], presented an EMA-based damage detection approach based on Principal Component Analysis (PCA) incorporated with ANNs. In [37], a new approach that utilizes a 1-D CNN has been introduced for detecting the general condition of a structure. This approach only requires two states of damage during the training stage, specifically undamaged and fully-damaged cases. The advantages in using 1-D CNNs in detecting structural damages were already inspected by the same authors in [38] and [39], where real-time capabilities of CNNs in detecting damages emerged. Shao et al. in [40] introduced a framework that utilizes Transfer Learning in a DL-based system for fault diagnosis. This approach enables and speeds up the training process of DNNs. Ai et al. in [41] proposed a novel approach based on 2D-CNNs for the raw EMA-based rapid damage quantification on structures. Tian et al. in [42] Bidirectional Long Short-Term Memory (LSTM) models to correlate girder vertical deflection and cable tension for condition assessment in SHM.

In [43], the authors proposed a DL framework that utilizes cloud computing to achieve efficient real-time monitoring and proactive maintenance of civil infrastructures. Cheng et al. in [11] introduced a data-driven method for performing health monitoring on machines, which is based on Adaptive Kernel Spectral Clustering (AKSC) and LSTM. In [44], a supervised anomaly detection method has been proposed by the authors, which utilizes a cluster of DNNs trained on time series signals transformed as grayscale images using computer vision techniques. In particular, in [44], clusters of DNNs are composed by stacked AEs trained by and greedy layer-wise training [45]. In [46], the authors presented an anomaly detection method that utilizes a Deep Coupling Autoencoder (DCAE) for handling multimodal sensory signals. The proposed method also integrates feature extraction of multimodal data into data fusion for fault diagnosis.

According to the growing interest in using AEs to solve general anomaly detection problems, several methods based on AEs for SHM damage-detection systems were proposed in literature. In [47], a monitoring method based on Conditional Convolutional AEs for identifying wind turbine blade breakages is proposed. Pathirage et al. in [48], [49], and [50] proposed several AE-based frameworks to learn the relationship between the physical properties of a structure and its vibration characteristics. The frameworks considered modal properties as input data and produced elemental stiffness reduction parameters of the structure as output. This was done to enable the detection of damages. In [51], a method based on DAE is proposed to extract damage features from data of undamaged structures affected by noise and temperature uncertainties. Mao et al. in [52] combine Generative Adversarial Networks (GAN) with AE to perform unsupervised damage classification on time series data that is transformed into images through Gramian Angular Field imaging. In [53], stacked AEs were used to extract damage-sensitive features from modal parameters of vibration raw data. Rastin et al. in [54] proposed convolutional AE to perform unsupervised damage detection on benchmark datasets leveraging on reconstruction error of AE. In [23], an unsupervised method based on acceleration signals was proposed. The method involved preprocessing the raw signals through Continuous Wavelet Transformation (CWT) and Fast Fourier Transformation (FFT), before feeding the data from each sensor into an AE to extract features. The extracted features were then classified as damaged or undamaged using an OC-SVM. The same authors in [55] proposed a novel method to detect, in an unsupervised manner, structural damages directly from raw acceleration responses (thus avoiding the use of CWT and FFT) using a OC-SVM fitted on damage-sensitive features extracted from original signals and their reconstruction made by the AE. Li et al. in [56] proposed a novel approach, the New Generalized Autoencoder (NGAE), which incorporates a statistical-pattern-recognition-based approach that lever-

ages on power cepstral coefficients of structural acceleration responses as damage-sensitive features to assess structural damages. In [57], Yan et al. presented a multi-domain indicator-based optimized stacked DAE to perform fault identification of rolling bearing.

However, a standard AE performs a deterministic mapping from the input data to its reconstruction, implying a lack in modeling data variability in latent representations [14]. This aspect involves several weaknesses in using an AE for anomaly detection tasks rather than a VAE, whose probabilistic encoder models the distribution parameters of the latent variables rather than the latent variables themselves [14], thus capturing more data variability and resulting in a more *homogeneous* latent space than a standard AE. The authors of [58] propose a novel anomaly detection approach that utilizes a combination of VAE and Support Vector Data Description (SVDD) [59]. In this approach, the SVDD decision boundary is learned simultaneously with the latent representations of data and fitted on them. This is done to prevent the problem of *hypersphere collapse*, which occurs when all the data points are mapped to a single point in the latent space [60]. Ma et al. presented a method based on VAEs in [61] to detect structural damages in the time-domain for SHM applications. The approach utilizes the latent representation obtained from the VAE's encoder to generate a time series of damage indexes during testing, which allows for the clear visualization of sudden changes in damage location. A method proposed in [62] employs a Convolutional VAE to extract features and performs anomaly detection using OC-SVM and Elliptic Envelope [63] on the learned latent representations. The authors of [64] proposed a damage detection approach that utilizes a VAE ensemble to calculate damage statistics based on Evidence Variational Lower Bound (ELBO) values. The ELBO values are then used to classify each input as damaged or undamaged using a decision rule defined by the user as a fixed threshold value. The authors of [65] proposed an unsupervised method for detecting tunnel damages from vibration data. The method uses a Convolutional VAE as a feature extractor and Wavelet Packet Decomposition (WPD) [66] to process the data and produce a damage index. The damage index is then compared to a fixed threshold value to classify the input data as damaged or undamaged. In [67] the authors proposed the Deep Order-Wavelet Convolutional Variational Autoencoder (DOWCVAE), a novel method for the identification of faults under fluctuating speed conditions. Xu et al. in [68] proposed a method based on VAE and GAN to assess the conditions of cable-stayed bridges. Yan et al. in [69] presented DRVAE, a novel DL model based on VAE for fault diagnosis of rotor–bearing system.

The approach presented in this work leverages on the advantages in using a VAE for anomaly detection [14] to perform damage detection in an SHM system. Differently from other methods, our proposal takes advantage of the VAE's probabilistic aspects to enhance the damage-sensitive feature extraction rather than using data latent representations modeled by VAE to detect damages. In particular, our proposal exploits the VAE's capability to model the undamaged data distribution through its probabilistic encoder during the training stage, in order to emphasize damaged data with different distributions. In this way, the difference in distributions is captured by the VAE's probabilistic decoder, which reconstructs the data less accurately as much as the damage increase. Finally, a OC-SVM is fitted on damage-sensitive features extracted by input data and their reconstruction in order to classify data as damaged or not.

III. PROPOSED ARCHITECTURE

In this work we propose a framework to perform a semi-supervised damage detection using a VAE followed by a OC-SVM. The main aim of our proposal consists in identifying the presence of damages regardless their intensity, thus producing outcomes from the application of this framework that can be interpreted in terms of a binary classification response.

A supervised method for identifying structural damage requires labeled data during the training phase, which means data must be recorded both in the undamaged and damaged states of the structure. However, in a real case study, the available data is assumed to be undamaged during the training phase. Therefore, the use of data on the damaged structure is subordinated to the adoption of Finite Element (FE) numerical models of the structure, which can simulate potential damage conditions. It should be noted that, for existing structures, the FE model is based on simplifying assumptions that may not fully match the experimental behavior of the structure. Updating the FE model can improve the accuracy of the simulation (e.g. by calibrating the matrix of masses and stiffnesses of the structure), but this process is time-consuming and requires extensive analysis. The described procedure, which uses a semi-supervised approach, circumvents this issue by relying solely on undamaged data during the training stage to detect structural decay without utilizing FE numerical models.

According to its definition, training a VAE on undamaged data involves the approximation of their intractable true posterior through their latent representation. In [70], an anomaly is defined as an observation that differs from regular data that it is considered to be generated by a different mechanism. This definition induces to consider distinct true posterior between undamaged and damaged data. Leveraging on this aspect, different latent distributions are generated by the probabilistic encoder if data are heterogeneous (i.e. including both undamaged and damaged data), thus inducing the probabilistic decoder to an erroneous data reconstruction if latent distributions are different from that of the undamaged data. Then, after a feature extraction stage, data are fed into a OC-SVM in order to learn a decision boundary to separate undamaged data from damaged data, and thus to classify new input datapoints as damaged or not. A representation of the framework is shown in Figure 1. In the following subsections VAE and OC-SVM models are explained.

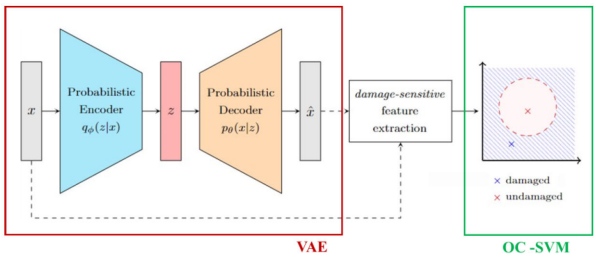


FIGURE 1. Graphical representation of the proposed architecture. Data are firstly fed into a VAE. Then, using original and reconstructed signals, after a feature extraction stage, data are fed into a OC-SVM for being classified as damaged or not.

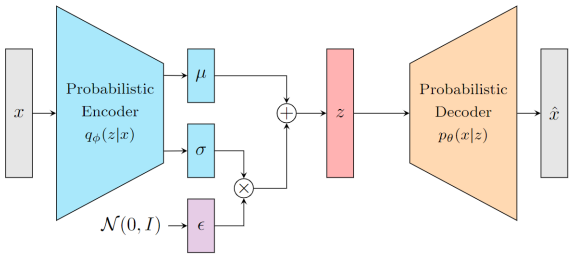


FIGURE 2. Architecture of a Variational Autoencoder.

A. VARIATIONAL AUTOENCODER

Considering x as data and z as its latent representation involved during the data generation process, a Variational Autoencoder (VAE) is a *probabilistic* generative model consisting of two main components: a probabilistic *decoder*, defined by a likelihood function $p_\theta(x|z)$, with parameters θ, that generates new data from a latent variable z, and a probabilistic *encoder*, defined by a posterior distribution $q_\phi(z|x)$, with parameters ϕ, that approximates the intractable true posterior $p_\theta(z|x)$.

To admit inference, VAE training simultaneously optimizes both the parameters θ and ϕ while learning the marginal likelihood of the data in the following generative process:

$$\max_{\phi,\theta} \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] \tag{1}$$

where $\log p_\theta(x|z)$ can be defined as:

$$\log p_\theta(x|z) = D_{KL}(q(z|x)||p(z)) + \mathcal{L}(\theta, \phi; x, z) \tag{2}$$

where $D_{KL}(\cdot)$ stands for the *Kullback–Leibler* (KL) divergence and $p(z)$ is the prior distribution over the latent variables z [71]. Notice that KL divergence quantifies the difference between two probability distributions q and p. Due to the non-negativity of the KL divergence, the term $\mathcal{L}(\theta, \phi; x, z)$ is called *Evidence Variational Lower Bound* (ELBO) on the marginal likelihood and it can be written as below:

$$\log p_\theta(x|z) \geq \mathcal{L}(\theta, \phi; x, z) = -D_{KL}(q_\phi(z|x)||p_\theta(z)) + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)]$$

where the second term is an *expected negative reconstruction error* between the input data and the data generated as output.

Leveraging on this formulation, VAE training can be performed by maximizing the ELBO [58]. However, the expected reconstruction error requires the sampling of random latent variables z from the approximated posterior $q_\phi(z|x)$, which makes the training intractable in practice since the gradient of the ELBO with respect to the parameters ϕ can not be estimated. This problem can be avoided using the *reparametrization trick*: assuming the prior $p(z)$ and the posterior $q_\phi(z|x)$ to be Gaussian distributions with a diagonal covariance matrix, with the prior $p(z)$ set to the isotropic unit Gaussian $\mathcal{N}(0, I)$, each random variable $z_i \sim q_\phi(z_i|x) = \mathcal{N}(\mu_i, \sigma_i)$ is reparametrized as differential transformation of a noise variable $\epsilon_i \sim \mathcal{N}(0, 1)$ as follows [71]:

$$z_i = \mu_i + \sigma_i \epsilon_i \tag{3}$$

Assuming the framework above, the ELBO can be differentiated and optimized with respect to both the variational parameters ϕ and θ [17]. In particular, ELBO can be maximized via gradient descent; this aspect involves a certain flexibility in modeling both the probabilistic encoder and the probabilistic decoder. A typical choice falls on the use of Multi-Layer Perceptron (MLP) Neural Networks [72]. In such case, the probabilistic encoder network takes the data x as input and computes the mean and the standard deviation of the approximate posterior $q_\phi(z|x)$ in order to sample the latent variable z. Then, the latent variable z is given as input of the decoder network which generates the reconstruction of the data $\hat{x}$. The architecture is shown in Figure 2.

B. ONE-CLASS SUPPORT VECTOR MACHINE

Considering input data as points defined in a vector space, a Support Vector Machine (SVM) [73] is a two-class method that classifies data according to a decision hyper-plane that maximizes the separation between the two classes. Researchers in SHM (Structural Health Monitoring) have been attracted by SVM due to its robustness in generalization capabilities [74], [75], [76]. However, in order to detect damages in a monitored structure, the use of a SVM implies that both of the undamaged and damaged data of the structure must be available during the training stage.

A One-Class Support Vector Machine (OC-SVM), instead, is a method that requires only data related to one class to train the model. The fundamental objective of the training stage in an OC-SVM is to determine a hyper-plane that can accurately define the region including the training samples [77]. This is achieved by solving the following optimization:

$$\begin{aligned} \min_{w,\xi_i,\rho} \quad & \frac{1}{2}\|w\|^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\xi_i - \rho \\ \textit{subject to} \quad & (w \cdot \Phi(x_i)) \geq \rho - \xi_i, \quad \xi_i \geq 0 \end{aligned} \tag{4}$$

where N refers to the number of training samples, w refers to the decision hyper-plane weights, x_i is the i-th training sample, $\Phi(\cdot)$ is a function that transforms data $\mathcal{X} \subseteq \mathbb{R}^d$ from

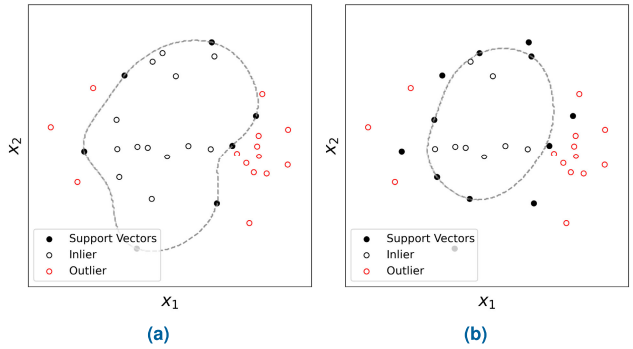


FIGURE 3. Graphical representation of an hyper-sphere fitted using a OC-SVM where $\nu = 0.1$ (a) and $\nu = 0.5$ (b) on data described by two features x_1 and x_2.

its original space into a new feature space $\mathcal{F} \subseteq \mathbb{R}^{d'}$ allowing the kernel trick $\Phi(x_i) \cdot \Phi(x_j) = K(x_i, x_j)$, ξ_i is a slack variable controlling how much error is allowed during the training stage and $\nu \in [0, 1]$ controls the proportion of outliers (i.e., training data lying outside the estimated region) as well as the number of support vectors.

Considering quadratic programming and Lagrange multipliers, the optimization problem above can be transformed into the following dual form:

$$\min_{\alpha} \quad \frac{1}{2} \sum_{i,j} \alpha_j \alpha_i K(x_j, x_i)$$
$$\textit{subject to} \quad 0 \leq \alpha_i \leq \frac{1}{N}, \quad \sum_{i=1}^{N} \alpha_i = 1 \tag{5}$$

where α_i is the Lagrange coefficient of the i-th training sample x_i. The non-zero coefficients α_i will determine the support vectors required to evaluate the decision function for a new test point x:

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i K(x, x_i) - \rho\right) \tag{6}$$

The test point x is outside the estimated region when the decision function $f(x)$ returns a negative value, otherwise it is inside [26], [55], [77]. In this work, we focus on the using of the Radial Basis Function (RBF) as the $\Phi(\cdot)$ function. In this way, the optimization problem involves the search of a hyper-sphere to estimate the region of the data rather than a hyper-plane. Moreover, we have set the parameter $\nu \approx 0$ since we are interested in capturing as many training samples as possible to determine the region of interest fitted by the OC-SVM. A graphical representation of a OC-SVM hyper-sphere is shown in Figure 3.

IV. EXPERIMENTAL ASSESSMENT

The architecture proposed in this work was evaluated on the benchmark dataset from the case study related to the steel frame tested in Phase II of the SHM benchmark problem [78], whose results were published in 2003 by the International Association for Structural Control (IASC) - American Society of Civil Engineers (ASCE) Structural Health Monitoring Task Group. The results of the experimental assessment are compared with the performances obtained by the method proposed in [37] on the same dataset and with the performances obtained by substituting VAE with a standard AE, thus following the approach proposed in [55]. In this Section, firstly details on the benchmark dataset are provided. Then, details regarding how data were arranged and specifics about the model selection stage involved in the experimental phase are described. Finally, results are shown and discussed.

FIGURE 4. Photo of the experimental setup [78].

A. CASE STUDY: EXPERIMENTAL PHASE II OF THE SHM BENCHMARK DATA

The frame is a four-story steel structure built at the University of British Columbia (Figure 4). The dimensions are 2.5 m × 2.5 m in plan, and the total height is 3.6 m. The structural elements are hot-rolled, grade 300W steel. The columns are B100 × 9 sections and beams are S75 × 11. In each span, the bracing system is composed of two threaded steel bars with a diameter of 12.7 mm and inserted along the diagonal. To make the mass distribution reasonably realistic, four slabs of 1000 kg are in the first, second and third floors, while slabs of 750 kg were used on the fourth. Further information can be read in [78].

Twelve accelerometers were placed on the structure as shown in Figure 5. On each floor, 3 accelerometers were installed on the west (in black), east (in red) and central column (in blue). All sensors are monoaxial: the accelerometers located on the west and on the east columns are oriented along the +X direction, while those on the central column are oriented along the +Y direction. In this paper, the signals are caused by shaker excitation, i.e., a band-limited white noise with components between 5–50 Hz.

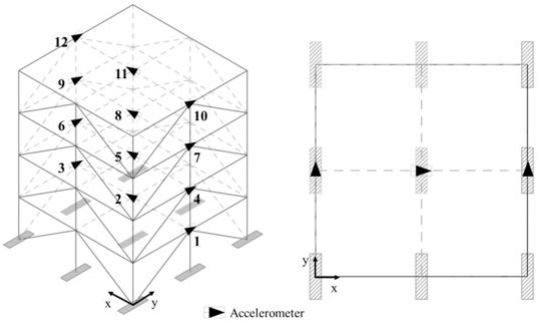


FIGURE 5. Location and direction of the sensors.

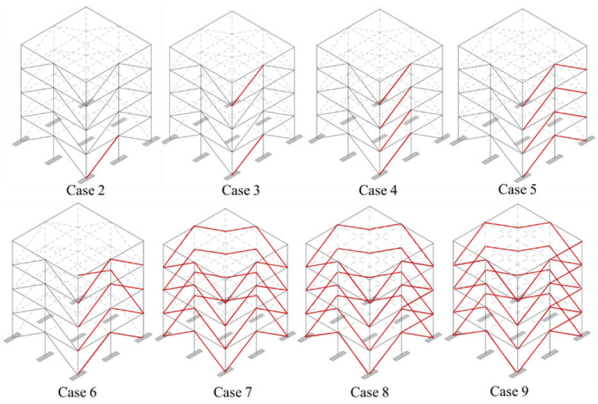


FIGURE 6. Damage scenarios.

Accelerations were recorded in the absence (Case 1) and in the presence of structural damage. Eight cases of damage were simulated. Table 1 and Figure 6 summarize the various damage scenarios in which the intensity gradually increases from Case 2 to Case 9. The simulated structural damage consists in the removal of diagonal stiffening elements in Cases 2 to 7, while the loosening of the connecting bolts is added in Cases 8 and 9. Figure 7 shows data distributions for each sensor and for each case.

B. DATA ARRANGEMENT

Data from Experimental Phase II were preprocessed following the setup proposed in [37]. In particular, each damage case S_i, with $1 \leq i \leq 9$, was considered as a set of signals collected by n sensors:

$$S_i = \{S_{i1}, S_{i2}, \ldots, S_{in}\}$$

Each signal S_{ij} of length d_j, with $1 \leq j \leq n$, was divided in a number of frames having the same length s:

$$S_{ij} = \{S_{ij,1}, S_{ij,2}, \ldots, S_{ij,n_{ij}}\}$$

where $n_{ij} = \lfloor d_j/s \rfloor$. Then, data were shuffled and normalized between 0 and 1, differently from [37] where data were normalized between -1 and 1. The normalization stage was performed considering minimum and maximum values computed through all the training dataset for each sensor. Before

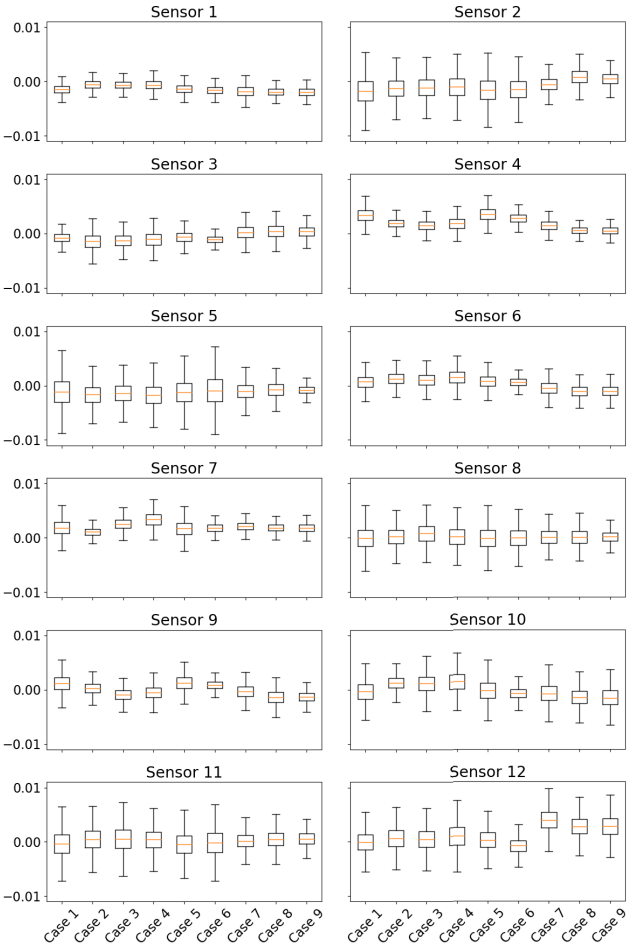


FIGURE 7. Graphical representation illustrating the data distributions for each sensor in every damage case of the benchmark dataset. Box plots were utilized to represent these distributions, and it can be observed that the distributions are not only significantly overlapping but also similar in the majority of the cases, suggesting that distinguishing between the damages may require additional analysis beyond examining the data alone.

starting the training stage, in order to have an estimate of the performances also on undamaged data, the 20 % of the samples from the Case 1 were extracted in order to evaluate the framework also on unseen undamaged data.

Following the experimental setup in [37], accelerations measured on the structure during the random shaker excitation under 5–50 Hz were used. Acceleration measurements were sampled at 200 Hz. Data were measured for 120 s for Cases 1 - 5, 300 s in Case 6 and for 360 s in the remaining cases. As it was explained above, an architecture for each accelerometer was trained using only undamaged data (Case 1). A length of $s = 128$ was considered to divide each signal in frames, thus obtaining 187 frames for Cases 1 - 5, 468 frames for Case 6 and 562 frames for Cases 7 - 9.

C. MODEL SELECTION

A fundamental phase in using machine learning algorithms consists in finding the best set of *hyperparameters*, i.e. the

TABLE 1. Structural cases description in the Phase II of the SHM benchmark problem [78].

Case	Description
1	Undamaged
2	On the first floor, diagonal element is removed in one bay
3	On the first and the fourth floors, diagonal elements are removed in one bay
4	On all floors, diagonal elements are removed in one bay
5	All braces are removed in the east face
6	On east face all braces are removed, while on north face of the second floor, braces are removed
7	All braces are removed
8	Case 7 + loosening of the connecting bolts for two beams
9	Case 7 + loosening of the connecting bolts for all beams in the east face

TABLE 2. Search spaces for bayesian optimization.

Module	Hyperparameter	Variation Range
VAE	N. of Layers	[1, 3], step: 1
	N. Neurons per Layer	[4, 128], step: 1
	Activation Function	{ReLU, LeakyReLU, Sigmoid}
	Latent dimension	[2, 40], step: 1
Training stage	Optimizer	{Adam, SGD}
	Learning Rate	{0.0001, 0.001, 0.01}

set of parameters of both the ML model and the learning algorithm which remain unchanged during the learning phase and whose values influence the final ML model performance on a given dataset [79]. This stage is often referred to as *model selection*. Examples of hyperparameters related to our proposal are the number of layers for the probabilistic encoder and the dimensionality of the latent space z of the VAE. Different approaches are known in literature to evaluate a ML model on some data during the hyperparameter search, such as the *holdout method* [80]. In our work, since only data related to the undamaged structure are involved in the training process, and since this set of data has a not-too-small number of samples, we chose *k-fold Cross-Validation*, that is commonly used for its statistical significance [79]. In particular, in our experiments we set $k = 10$ to determine the data partitioning. In order to explore and evaluate different sets of hyperparameters, we referred to *hyperparameter optimization* algorithms since, due to the high number of hyperparameters of the overall architecture, a manual tuning could have been too much expensive from a timing perspective. Among the different algorithms proposed in literature, our choice fell on the *bayesian optimization* [81].

In this work, VAE model selection stage was performed separately for each sensor considering 100 trials for the bayesian optimization in order to *minimize* the averaged reconstruction error on validation sets produced by the k-fold Cross-Validation. MLP Neural Networks were adopted as architecture to model both the probabilistic encoder and probabilistic decoder. Search spaces for hyperparameters were established during a preliminary manual analysis with the aim of minimizing the computational time needed for the overall model selection stage. The specific details of these search spaces can be found in Table 2. For each fold, the 20 % of the data were extracted from the training set and considered as validation set. The number of epochs was set to 1000 and the early stopping criterion was considered as convergence criterion with a patience of 50 epochs.

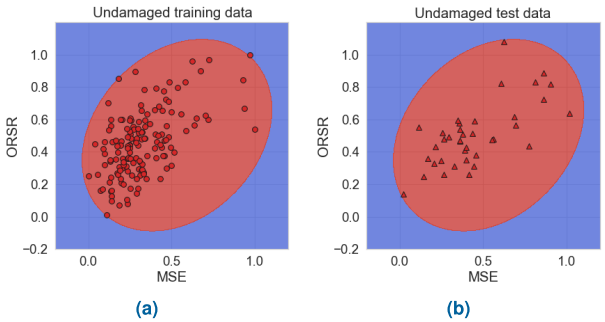


FIGURE 8. Graphical representation of a OC-SVM fitted on undamaged (Case 1) training data (a) and tested on undamaged testing data (b).

As a result of the model selection stages, Shallow Neural Networks (i.e., MLP Neural Network having 1 hidden layer) with the Sigmoid as activation function resulted to be the best architecture for VAE's probabilistic decoders and probabilistic encoders. Since the number of neurons in the hidden layer and the latent dimension assumed values respectively in neighborhoods of 40 and 20 reporting similar performances, we fixed the final configuration of each network as having 40 neurons in the hidden layer and 20 neurons for the latent representation. VAE's training stages were performed using Adam optimizer [82] with a learning rate of 0.001. The OC-SVM's parameter ν was fixed to 0.001 and the RBF was considered as kernel function. An example of undamaged region fitted by the OC-SVM is shown in Figure 8.

D. RESULTS

In this subsection, the experimental results related to the application of our proposal on the benchmark problem are reported. As in [55], the following damage-sensitive features were considered:

1) *Mean Squared Error* (MSE), which measures the reconstruction error between the input acceleration signals and their reconstruction as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2 \tag{7}$$

where n is the number of the signal features, x_i is the i-th feature in the original signal and $\hat{x}_i$ is the i-th feature in the reconstructed signal;

2) *Original-to-Reconstructed-Signal Ratio* (ORSR), computed as:

$$ORSR = 10 \log_{10} \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} \hat{x}_i^2} \tag{8}$$

that represents the ratio in decibels between the magnitudes of the original signal and its reconstruction.

The method performance evaluation was obtained by the score used in [37] in order to make a comparison of the results. Thus, to each set S_{ij}, the probability of damage (PoD) was computed as follows:

$$PoD_{ij} = \frac{c_{ij}}{n_{ij}} \times 100 \tag{9}$$

where c_{ij} is the number of samples classified as damaged by the OC-SVM. Finally, the overall structure score for each case S_i was computed by averaging the PoD values of each sensor:

$$PoD_{avg,i} = \frac{PoD_{i1} + PoD_{i2} + \ldots + PoD_{in}}{n} \tag{10}$$

As it was described in [37], a low value of PoD_{ij} indicates a low probability that the signal i recorded by the j-th sensor belongs to an undamaged state. On the other hand, a high value indicates a high probability of belonging to damaged state. Same observations are valid for the $PoD_{avg,i}$ value.

Experimental results are reported in Table 3. We remark that the main aim of our proposal consists in perform damage detection from data. The PoD values of each sensor are interpreted as the probability of belonging to the damaged state, considering a PoD value of 0 % as an undamaged structure, 100 % as a damaged structure and 50% as a chance probability.

We can notice that the PoD_{avg} values reflect the a priori known damage conditions of the structure: damage probability is low for Case 1 (i.e., undamaged case), while it is high for all the remaining cases (i.e., damaged cases). It is worth noticing that PoD_{avg} values higher than the $\sim$ 89% are always reached, except for Case 2 and Case 6, where PoD_{avg} values of $\sim$ 70% resulted as outcome. In Case 2, we can notice that the PoD_{avg} is decreased by the PoD values related to the central sensors. For each damaged case, PoD values of each sensor are not correlated to mutual position sensor-damage. Therefore, the choice to calibrate the framework for each sensor does not allow us to do damage localization.

Nevertheless, the proposed approach can suggest which are the most efficient sensors to be selected to monitor a structure (such as sensors 3, 4 and 12). For instance, Figure 9 shows that the damage is better detected by sensor 12 (lateral sensor) than sensor 2 (central sensor).

In [37], PoD_{avg} values related to Case 2 and Case 6 are estimated to be respectively $\sim$ 22% and $\sim$ 50%, while in our case they are estimated to be $\sim$ 70% and $\sim$ 71%. According to a probability perspective, results reported by [37] are close

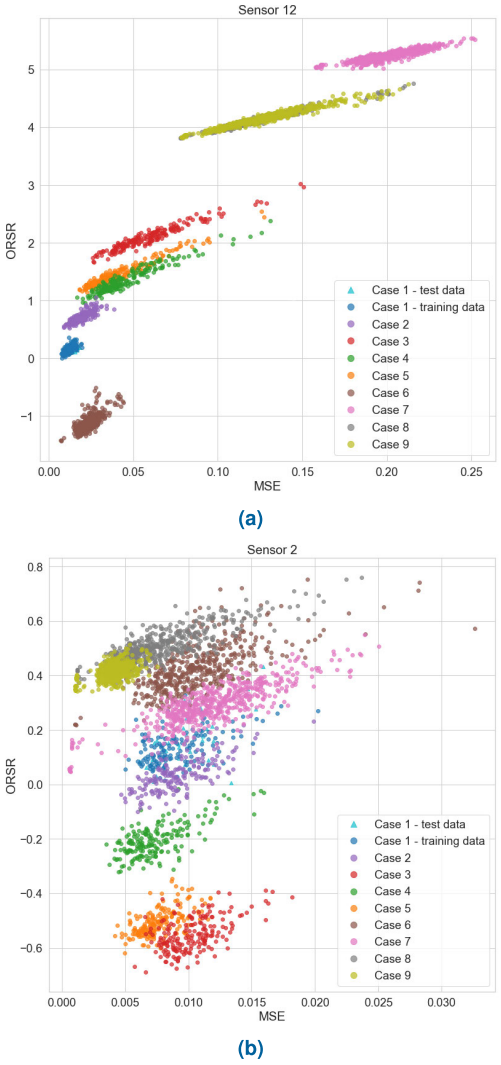


FIGURE 9. Graphical representation of damage-sensitive features extracted from sensor 12 (a) and sensor 2 (b).

to the chance probability for Case 6, and is close to an undamaged probability for Case 2, while in our case the presence of structural damages is suggested in both the cases. Similar observations can be done for the remaining cases shown in [37], such as Cases 3, 4 and 5, where PoD_{avg} values don't suggest the presence of a damage, even if present. Moreover, PoD_{avg} values in [37] hide PoD values close to 0 and 100, thus giving a not-too-reliable estimate of the overall structural conditions in some cases: for example, Case 4 is reported to have a PoD_{avg} value of 39.77 ± 36.24, having $min = 0$ and $max = 100$ suggesting, respectively, a fully undamaged and damaged condition of the structure; in our case instead, Case 4 is reported to have a PoD_{avg} value of 99.96 ± 0.16, having $min = 96.47$ and $max = 100$, thus reporting a more reliable summary of the structural condition.

It is also important to point out that, differently from [37] where a supervised damage detection method was proposed, we propose a semi-supervised methodology for damage

TABLE 3. Results on the nine structural cases. For each sensor (rows 1-12), after a description regarding the sensor position (columns 1-3), *PoD* values are reported for all the Cases (columns 4-12). The last row reports the *PoD* values averaged for each Case. In parenthesis, the difference from the results using a standard AE is reported.

ID	Location	Orientation	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
1	1st Floor / West	N/S	0 (5.26)	99.47 (0.53)	100 (0)	100 (0)	100 (0)	98.72 (0.64)	100 (0)	100 (0)	100 (0)
2	1st Floor / Center	E/W	5.26 (-5.26)	8.08 (2.62)	100 (0)	100 (0)	100 (-0.53)	72.65 (-30.98)	100 (0)	100 (0)	100 (0)
3	1st Floor / East	N/S	15.79 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)
4	2nd Floor / West	N/S	2.63 (24.05)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)
5	2nd Floor / Center	E/W	7.89 (34.21)	37.97 (-5.88)	100 (0)	99.47 (0.53)	100 (0)	96.37 (0.85)	33.10 (61.92)	100 (0)	100 (0)
6	2nd Floor / East	N/S	10.53 (-7.90)	98.93 (-75.94)	100 (0)	100 (0)	100 (0)	13.25 (-1.28)	100 (0)	100 (0)	100 (0)
7	3rd Floor / West	N/S	0 (18.42)	96.79 (2.68)	100 (0)	100 (0)	100 (0)	10.26 (-1.28)	100 (0)	12.28 (0)	37.72 (0)
8	3rd Floor / Center	E/W	2.63 (0)	4.28 (-1.07)	97.86 (-31.55)	100 (0)	96.26 (-39.58)	13.46 (-6.41)	58.90 (-16.37)	52.31 (-30.60)	99.47 (-19.58)
9	3rd Floor / East	N/S	0 (2.63)	98.93 (-2.67)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)
10	4th Floor / West	N/S	0 (26.32)	81.28 (7.49)	100 (0)	100 (0)	100 (0)	100 (0)	99.82 (0.18)	100 (0)	100 (0)
11	4th Floor / Center	E/W	10.53 (13.15)	17.11 (0.54)	100 (0)	100 (0)	100 (0)	46.58 (-22.01)	100 (0)	100 (0)	100 (0)
12	4th Floor / East	N/S	0 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)	100 (0)
$\mathrm{PoD_{avg}}$			**4.61 (8.99)**	**70.23 (-5.97)**	**99.82 (-2.63)**	**99.96 (0.04)**	**99.69 (-3.34)**	**70.94 (-5.20)**	**90.98 (3.82)**	**88.71 (-1.74)**	**94.77 (0.80)**

detection, where only undamaged data are necessary for the training stage.

V. ANALYSIS ON THE IMPACT OF THE VAE

Differently from [55], where damage detection is performed using an architecture composed by an AE followed by a OC-SVM, in our proposal anomaly detection is performed using a VAE followed by a OC-SVM. As in [55], data, before being fed as input to the OC-SVM, are transformed using damage-sensitive features extracted from the original signals and their reconstruction made by VAE. As we have described above, a VAE has the capability of learning to produce distributions of data through latent representations generated by its probabilistic encoder. Moreover, differently from standard AEs, VAEs don't learn a deterministic mapping from input to their reconstruction, thus modeling data variability in latent representations [14]. In order to verify the advantages of using a VAE instead of an AE on the proposed method, an experimental assessment was made substituting VAE with a standard AE while maintaining the same architectures. Results are shown in Table 3 in parenthesis as difference from the results obtained through the use of VAE. We can observe that the PoD_{avg} value related to the undamaged case (Case 1) is higher than the one reached by our proposal, thus exhibiting a lower capability in recognizing undamaged data than our architecture. Moreover, we can notice that PoD_{avg} values for almost all the cases are lower than those reached by our proposal, involving that damages are detected with lower probabilities than our architecture. This aspect implies that the use of a VAE entails a more robust damage probability estimation than using a standard AE (4.65% improvement on average). A graphical representation of the PoD_{avg} obtained through VAE and AE is reported in Figure 10.

Assuming that generating distributions of damaged data are different from that of undamaged data, our proposal aims to learn the latent distribution of undamaged data in order to induce the probabilistic encoder to encode damaged data with different generating distributions. As a consequence, the probabilistic decoder will hardly decode data coming from distributions diverse from those learned during the training stage, thus resulting in high reconstruction error. In order to verify how much generating distributions of damaged data diverge from that of undamaged data, KL divergences were

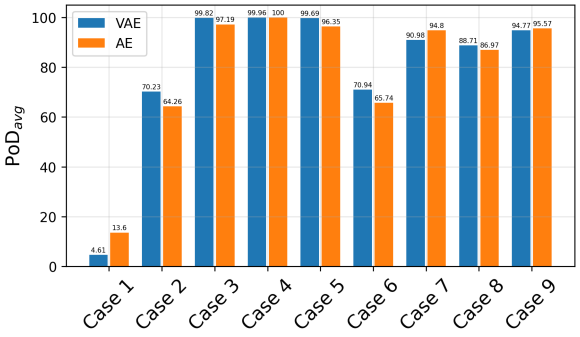


FIGURE 10. Graphical comparison of PoD_{avg} values obtained using VAE (blue) and AE (orange).

TABLE 4. For each sensor, the KL divergences of damaged cases from the undamaged case (Case 1) is shown. On the last row, the averaged KL divergence is represented for each case.

Sensor ID	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
1	0.069	0.068	0.088	0.085	0.210	0.225	0.214	0.215
2	0.048	0.048	0.050	0.056	0.172	0.192	0.192	0.192
3	0.063	0.066	0.088	0.075	0.183	0.205	0.214	0.204
4	0.043	0.046	0.058	0.052	0.166	0.203	0.191	0.196
5	0.049	0.049	0.051	0.056	0.180	0.193	0.193	0.193
6	0.041	0.038	0.040	0.042	0.160	0.192	0.185	0.187
7	0.039	0.040	0.042	0.047	0.162	0.187	0.187	0.187
8	0.065	0.066	0.069	0.073	0.192	0.208	0.215	0.205
9	0.044	0.046	0.056	0.052	0.167	0.192	0.192	0.192
10	0.051	0.052	0.057	0.063	0.176	0.198	0.197	0.197
11	0.053	0.055	0.055	0.062	0.181	0.195	0.196	0.196
12	0.045	0.045	0.046	0.055	0.172	0.192	0.191	0.192
$\mathrm{KL_{avg}}$	**0.051**	**0.052**	**0.058**	**0.060**	**0.177**	**0.198**	**0.197**	**0.196**

computed for each sensor and reported in Table 4. Recall that KL divergence quantifies the difference between two probability distributions q and p. We can notice from the averaged KL values reported as KL_{avg} in Table 4 that latent distributions of damaged data diverge as much as damages increase, thus confirming the assumptions made above. This aspect suggests that latent representations become harder to decode by the probabilistic decoder of VAE as the damages increase (Figure 11). Moreover, the increasing damages captured by VAE's approximation of generating distributions implies that the amount of damages is implicitly suggested in the damage identification process of our architecture. Using t-SNE [83], latent representations of each case related to a randomly chosen sensor are shown in Figure 12.

A traditional method for damage identification in structures is the Frequency Domain Decomposition (FDD) [18]. The method allows identifying the frequencies associated

TABLE 5. FDD results.

Mode	Frequencies (Hz)								
	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
1	7.47 (0)	7.47 (0)	7.32 (-2)	6.64 (-11.11)	5.18 (-30.66)	5.96 (-20.21)	2.63 (-64.79)	2.54 (-66)	2.58 (-65.46)
2	7.76 (0)	-	7.46 (3.9)	7.62 (1.80)	7.71 (-0.64)	7.81 (0.64)	3.62 (-53.35)	3.28 (-57.73)	3.37 (-56.57)

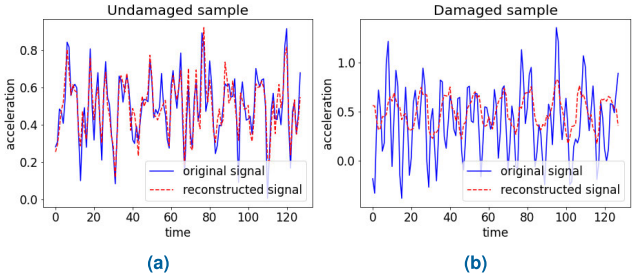


FIGURE 11. Graphical representation of an undamaged (i.e., Case 1) (a) and a damaged (i.e., Case 7) (b) signal reconstructed by a VAE.

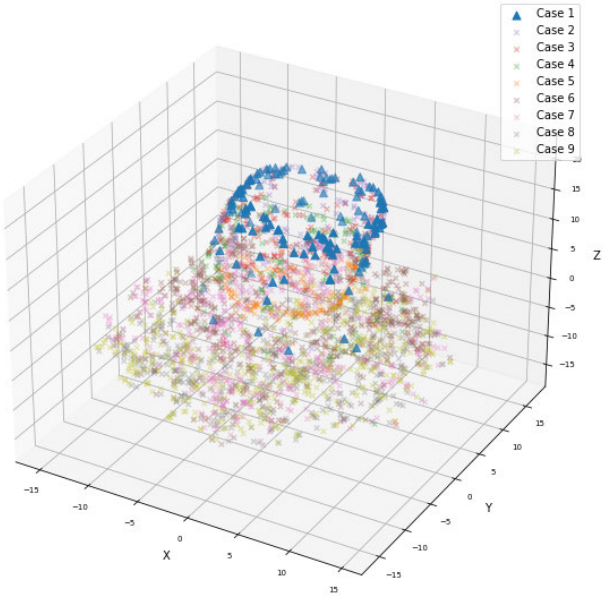


FIGURE 12. Graphical representation of latent representations for each case using t-SNE.

with the vibration modes of a structure based on the analysis of the accelerations recorded on the structure, due to natural vibration or shaking. A change in frequency indicates a change in stiffness: if the frequency decreases, the structure is more deformable and this could indicate that the structure is experiencing damage.

Table 6 shows the frequencies of the first two vibration modes of the healthy structure (Case 1) and the eight damaged structures (Case 2 - 9), obtained by FDD. Variation in percentage for each damaged case from the undamaged case is shown in brackets. The traditional FDD technique is scarcely able to detect damages for Case 2 due to low damage intensity, while it is able to detect damages for Cases 7, 8, and 9 where the frequency values decrease significantly (more than 60%) because they are characterized by the presence of several ''damaged'' elements. On the contrary, our method identifies all the different structural conditions.

Finally, by comparing the variations in percentage shown in Table 5 with the KL_{avg} values listed in Table 4, we can notice a correspondence between the KL values obtained through the DL-based method and the frequency variations obtained through traditional FDD method: higher the frequency variation, higher the KL value. Thus, we could consider the KL value as a parameter suggesting a quantification of the damage, differently from [37] where the *PoD* values were considered to estimate the quantification of damage.

VI. NOISE IMPACT ANALYSIS

A series of experiments was conducted to assess the performance of the proposed method across various simulated noise scenarios. Gaussian noise with different sigma levels was introduced to simulate the noise conditions. Since the input signal's magnitude was on the order of 10^{-3}, the sigma level was gradually increased until it reached this threshold.

Figure 13 shows the effect of increasing noise factors on the data in two different scenarios, i.e. when noise is already during the training stage (a) and when noise emerges over time following the completion of the training stage (b). We can notice that the presence of noise alters the performances of the proposed pipeline only when its level reaches a magnitude comparable to that of the signal data (i.e., 10^{-3}), thus revealing that the pipeline is resistant to noise level either when it is already present during the training stage or when it occurs over time.

The traditional technique based on dynamic identification is not effective when the data are influenced by noise. In particular, the representation of the first singular value of the power spectrum is strongly distorted by noise when sigma is between 10^6 to 10^3. Indeed, the resonance peaks - from which the vibration eigenfrequency of the structure can be read - are not detected. Conversely, when the noise is reduced, the frequencies are uniquely determined.

Figure [...] shows the representation of the first singular value of the decomposed spectrum. The curve for the case without noise (i.e., when data are filtered) is presented in black. The other colors represent the curves obtained with raw data by adding noise. Therefore, frequency variation used as a damage-sensitive feature - and consequently, the traditional method - are inefficient in the presence of noise because the latter affects the detection of the frequencies themselves i.e., it does not allow their identification.

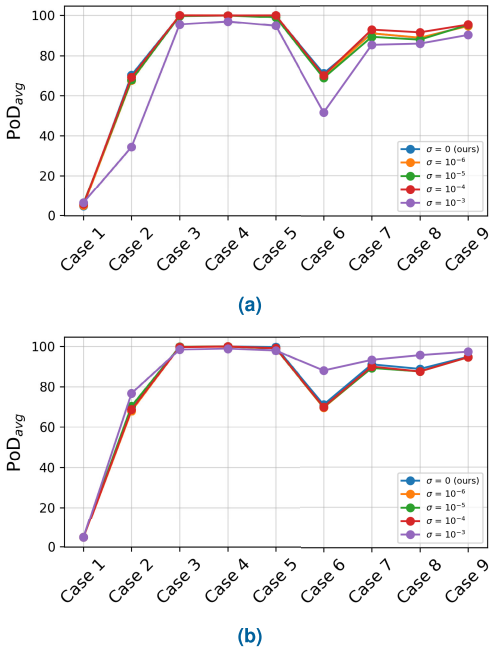


FIGURE 13. An investigation into the influence of noise factors in two distinct scenarios: when noise is initially present during the training stage (a); when noise emerges over time following the completion of the training stage (b).

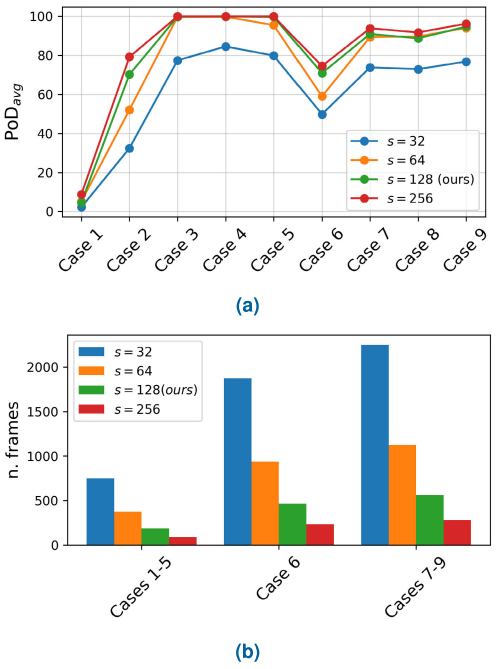


FIGURE 15. An investigation into the influence of the frame size s on both our proposed method (a) and the numerosity of the dataset (b).

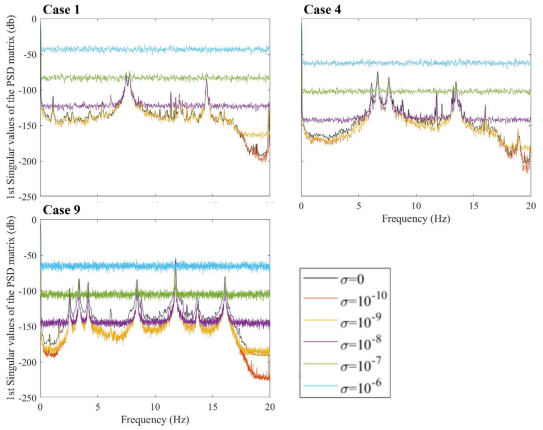


FIGURE 14. The influence of noise on the representation of the first singular value of decomposed PSD for healthy (1) and damaged cases (4 and 9).

VII. REMARKS

In this work, we proposed a framework to perform a semi-supervised damage detection in an SHM system based on a VAE and a OC-SVM in order to minimize human interactions during the data classification process. It is important to note that, even though we have focused our studies on MLP, VAEs can be implemented using various other architectures, such as CNNs and RNNs. While we acknowledge that different implementations of VAEs can potentially impact the overall performance of the pipeline, our study primarily focused on examining the functionality of the entire framework to gain insights into its operation. Moreover, it is worth mentioning that there exist alternative generative methods for anomaly detection that could also be explored, e.g. GANs. Additionally, among other ML approaches such as SVDD or clustering algorithms that may also provide valuable insights, we focused on OC-SVM since it defines a decision boundary and offers advantages such as providing a good control over its definition through several hyperparameters.

Moreover, we have implemented $s = 128$ in accordance with the setup proposed by [37] as stated above. However, it is essential to highlight that the dimensionality of the sample could yield different outcomes. Figure 15 demonstrates that a sample size lower than ours may result in reduced information contained in the samples, leading to lower PoD_{avg}, despite an increase in the number of samples. Conversely, incorporating more context (such as $s = 256$) can improve accuracy, even with a decrease in the number of samples. It is worth noting that despite this consideration, $s = 128$ appears to be a favorable compromise, as its performance closely aligns with that of 256. Thus, it is plausible that achieving the same result may be possible with a larger sample size.

Finally, for Case 6, certain sensors (specifically sensors 6, 7, and 8) fail in detecting the presence of damage, whereas the remaining sensors exhibit high PoD values. Despite the PoD_{avg} value being reasonably high (approximately 70%), this outcome highlights two aspects. Firstly, there is room for improvement in the algorithm to better identify minor anomalies in the measurements obtained from less damage-sensitive sensors. Secondly, it is important to note that relying solely on the PoD_{avg} value derived from trained networks for each sensor could lead to inaccuracies when numerous sensors lack sensitivity to damage.

VIII. CONCLUSION

In this work, we proposed a framework that allows to automate the entire damage identification process (from the training stage to the testing stage) requiring less time than a traditional SHM technique. In particular, if we consider a typical SHM technique (i.e. FDD) that compares the frequency of vibration of the structural system in different conditions to identify anomalies, we have to highlight that (i) the frequency identification is not always unique (ii) the threshold to define if there is an anomaly is completely arbitrary.

The probabilistic aspects of a VAEs allow to model data heterogeneity with different generating distributions. In the case of undamaged/damaged data, the probabilistic encoder models different data distribution thus involving an implicit capture of damaged states of a structure and resulting in a more robust damage-detection system than using a standard AE. Moreover, the KL divergence, which is generally implied in VAE's training stage, could be evaluated for the cases in which a damage is detected in order to quantify it.

Currently, as we have seen in the discussion of the experimental assessment, our framework does not give the possibility to localize a damage according to the score obtained by the single sensors. Recently, several methods were proposed to interpret decisions of anomaly detection methods using XAI techniques [84]. For this reason, in future works, we would like to extend our framework in order to give the possibility not only to detect general damages of the structure, but also to reliably identify where the damages are located. Moreover, in future works, we aim to extend the application of our methodology to more complex structures associated with real-life case studies. This will enable us to evaluate the efficacy and robustness of our approach in practical real-world scenarios. In this scenario, we intend to tackle scenarios where the normal condition of a structure deviates from its established normal state, outlined in the training data, through a new normal condition. Novel normal state could be determined by several causes. such as changing loads. In this case, we would explore possibilities for adapting the existing normal state to accommodate the new conditions through a refined learning process, such as Transfer Learning techniques.

REFERENCES

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 1–58, 2009.

[2] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *J. Netw. Comput. Appl.*, vol. 60, pp. 19–31, Jan. 2016.

[3] M. Canizo, I. Triguero, A. Conde, and E. Onieva, "Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study," *Neurocomputing*, vol. 363, pp. 246–260, Oct. 2019.

[4] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019, *arXiv:1901.03407*.

[5] S. Omar, A. Ngadi, and H. H. Jebur, "Machine learning techniques for anomaly detection: An overview," *Int. J. Comput. Appl.*, vol. 79, no. 2, pp. 33–41, Oct. 2013.

[6] G. Pang, C. Shen, L. Cao, and A. V. D. Hengel, "Deep learning for anomaly detection: A review," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–38, 2021.

[7] H. Liang, L. Song, J. Wang, L. Guo, X. Li, and J. Liang, "Robust unsupervised anomaly detection via multi-time scale DCGANs with forgetting mechanism for industrial multivariate time series," *Neurocomputing*, vol. 423, pp. 444–462, Jan. 2021.

[8] N. Li and F. Chang, "Video anomaly detection and localization via multivariate Gaussian fully convolution adversarial autoencoder," *Neurocomputing*, vol. 369, pp. 92–105, Dec. 2019.

[9] J. Fan, Q. Zhang, J. Zhu, M. Zhang, Z. Yang, and H. Cao, "Robust deep auto-encoding Gaussian process regression for unsupervised anomaly detection," *Neurocomputing*, vol. 376, pp. 180–190, Feb. 2020.

[10] C. Zhou and R. C. Paffenroth, "Anomaly detection with robust deep autoencoders," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 665–674.

[11] Z. Chen, C. K. Yeo, B. S. Lee, and C. T. Lau, "Autoencoder-based network anomaly detection," in *Proc. Wireless Telecommun. Symp. (WTS)*, Apr. 2018, pp. 1–5.

[12] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proc. 2nd Workshop Mach. Learn. Sensory Data Anal.*, Dec. 2014, pp. 4–11.

[13] J. K. Chow, Z. Su, J. Wu, P. S. Tan, X. Mao, and Y. H. Wang, "Anomaly detection of defects on concrete structures with the convolutional autoencoder," *Adv. Eng. Informat.*, vol. 45, Aug. 2020, Art. no. 101105.

[14] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special Lecture IE*, vol. 2, pp. 1–18, Dec. 2015.

[15] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 1096–1103.

[16] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8595–8598.

[17] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[18] R. Brincker, L. Zhang, and P. Andersen, "Modal identification of output-only systems using frequency domain decomposition," *Smart Mater. Struct.*, vol. 10, no. 3, pp. 441–445, Jun. 2001.

[19] J. S. Kang, S.-K. Park, S. Shin, and H. S. Lee, "Structural system identification in time domain using measured acceleration," *J. Sound Vibrat.*, vol. 288, nos. 1–2, pp. 215–234, Nov. 2005.

[20] A. Alvandi and C. Cremona, "Assessment of vibration-based damage identification techniques," *J. Sound Vibrat.*, vol. 292, nos. 1–2, pp. 179–202, Apr. 2006.

[21] C. R. Farrar, S. W. Doebling, and D. A. Nix, "Vibration-based structural damage identification," *Philos. Trans. Roy. Soc. Lond. A, Math., Phys. Eng. Sci.*, vol. 359, no. 1778, pp. 131–149, Jan. 2001.

[22] H. Zhu, H. Yu, F. Gao, S. Weng, Y. Sun, and Q. Hu, "Damage identification using time series analysis and sparse regularization," *Struct. Control Health Monitor.*, vol. 27, no. 9, p. e2554, Sep. 2020.

[23] Z. Wang and Y.-J. Cha, "Automated damage-sensitive feature extraction using unsupervised convolutional neural networks," *Proc. SPIE*, vol. 10598, Mar. 2018, Art. no. 105981J.

[24] H. Li, J. Ou, X. Zhao, W. Zhou, H. Li, Z. Zhou, and Y. Yang, "Structural health monitoring system for the Shandong Binzhou Yellow River Highway Bridge," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 21, no. 4, pp. 306–317, May 2006.

[25] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017.

[26] J. Long and O. Buyukozturk, "Automated structural damage detection using one-class machine learning," in *Dynamics of Civil Structures*, vol. 4. Berlin, Germany: Springer, 2014, pp. 117–128.

[27] M. Azimi, A. Eslamlou, and G. Pekcan, "Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review," *Sensors*, vol. 20, no. 10, p. 2778, May 2020.

[28] Z. Hong-ping, H. Bo, and C. Xiao-qiang, "Detection of structural damage through changes in frequency," *Wuhan Univ. J. Natural Sci.*, vol. 10, no. 6, pp. 1069–1073, Nov. 2005.

[29] E. P. Carden and P. Fanning, "Vibration based condition monitoring: A review," *Struct. Health Monitor.*, vol. 3, no. 4, pp. 355–377, Dec. 2004.

[30] X. Yan, Y. Liu, and M. Jia, "Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions," *Knowl.-Based Syst.*, vol. 193, Apr. 2020, Art. no. 105484.

[31] X. Yan, Y. Liu, Y. Xu, and M. Jia, "Multistep forecasting for diurnal wind speed based on hybrid deep learning model with improved singular spectrum decomposition," *Energy Convers. Manag.*, vol. 225, Dec. 2020, Art. no. 113456.

[32] Y. Xu, W. Qian, N. Li, and H. Li, "Typical advances of artificial intelligence in civil engineering," *Adv. Struct. Eng.*, vol. 25, no. 16, pp. 3405–3424, Dec. 2022.

[33] S. Li, X. Zuo, Z. Li, and H. Wang, "Applying deep learning to continuous bridge deflection detected by fiber optic gyroscope for damage detection," *Sensors*, vol. 20, no. 3, p. 911, Feb. 2020.

[34] H. Li, D. Ai, H. Zhu, and H. Luo, "Integrated electromechanical impedance technique with convolutional neural network for concrete structural damage quantification under varied temperatures," *Mech. Syst. Signal Process.*, vol. 152, May 2021, Art. no. 107467.

[35] D. Ai, F. Mo, Y. Han, and J. Wen, "Automated identification of compressive stress and damage in concrete specimen using convolutional neural network learned electromechanical admittance," *Eng. Struct.*, vol. 259, May 2022, Art. no. 114176.

[36] D. Ai, F. Mo, F. Yang, and H. Zhu, "Electromechanical impedance-based concrete structural damage detection using principal component analysis incorporated with neural network," *J. Intell. Mater. Syst. Struct.*, vol. 33, no. 17, pp. 2241–2256, Oct. 2022.

[37] O. Abdeljaber, O. Avci, M. S. Kiranyaz, B. Boashash, H. Sodano, and D. J. Inman, "1-D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data," *Neurocomputing*, vol. 275, pp. 1308–1317, Jan. 2018.

[38] O. Avci, O. Abdeljaber, S. Kiranyaz, and D. Inman, "Structural damage detection in real time: Implementation of 1D convolutional neural networks for SHM applications," in *Structural Health Monitoring & Damage Detection*, vol. 7. Berlin, Germany: Springer, 2017, pp. 49–54.

[39] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, and D. J. Inman, "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks," *J. Sound Vibrat.*, vol. 388, pp. 154–170, Feb. 2017.

[40] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.

[41] D. Ai and J. Cheng, "A deep learning approach for electromechanical impedance based concrete structural damage quantification using two-dimensional convolutional neural network," *Mech. Syst. Signal Process.*, vol. 183, Jan. 2023, Art. no. 109634.

[42] Y. Tian, Y. Xu, D. Zhang, and H. Li, "Relationship modeling between vehicle-induced girder vertical deflection and cable tension by BiLSTM using field monitoring data of a cable-stayed bridge," *Struct. Control Health Monitor.*, vol. 28, no. 2, p. e2667, Feb. 2021.

[43] H. V. Dang, M. Tatipamula, and H. X. Nguyen, "Cloud-based digital twinning for structural health monitoring using deep learning," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 3820–3830, Jun. 2022.

[44] Y. Bao, Z. Tang, H. Li, and Y. Zhang, "Computer vision and deep learning-based data anomaly detection method for structural health monitoring," *Struct. Health Monitor.*, vol. 18, no. 2, pp. 401–421, Mar. 2019.

[45] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, "Greedy layer-wise training of deep networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 19, 2006, pp. 1–8.

[46] M. Ma, C. Sun, and X. Chen, "Deep coupling autoencoder for fault diagnosis with multimodal sensory data," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1137–1145, Mar. 2018.

[47] L. Yang and Z. Zhang, "A conditional convolutional autoencoder-based method for monitoring wind turbine blade breakages," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6390–6398, Sep. 2021.

[48] C. S. N. Pathirage, J. Li, L. Li, H. Hao, and W. Liu, "Application of deep autoencoder model for structural condition monitoring," *J. Syst. Eng. Electron.*, vol. 29, no. 4, pp. 873–880, Aug. 2018.

[49] C. S. N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, and P. Ni, "Structural damage identification based on autoencoder neural networks and deep learning," *Eng. Struct.*, vol. 172, pp. 13–28, Oct. 2018.

[50] C. S. N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, and R. Wang, "Development and application of a deep learning-based sparse autoencoder framework for structural damage identification," *Struct. Health Monitor.*, vol. 18, no. 1, pp. 103–122, Jan. 2019.

[51] Z. Shang, L. Sun, Y. Xia, and W. Zhang, "Vibration-based damage detection for bridges by deep convolutional denoising autoencoder," *Struct. Health Monitor.*, vol. 20, no. 4, pp. 1880–1903, Jul. 2021.

[52] J. Mao, H. Wang, and B. F. Spencer, "Toward data anomaly detection for automated structural health monitoring: Exploiting generative adversarial nets and autoencoders," *Struct. Health Monitor.*, vol. 20, no. 4, pp. 1609–1626, Jul. 2021.

[53] M. F. Silva, A. Santos, R. Santos, E. Figueiredo, and J. C. W. A. Costa, "Damage-sensitive feature extraction with stacked autoencoders for unsupervised damage detection," *Struct. Control Health Monitor.*, vol. 28, no. 5, p. e2714, May 2021.

[54] Z. Rastin, G. G. Amiri, and E. Darvishan, "Unsupervised structural damage detection technique based on a deep convolutional autoencoder," *Shock Vibrat.*, vol. 2021, pp. 1–11, Apr. 2021.

[55] Z. Wang and Y.-J. Cha, "Unsupervised deep learning approach using a deep auto-encoder with a one-class support vector machine to detect damage," *Struct. Health Monitor.*, vol. 20, no. 1, pp. 406–425, Jan. 2021.

[56] L. Li, M. Morgantini, and R. Betti, "Structural damage assessment through a new generalized autoencoder with features in the quefrency domain," *Mech. Syst. Signal Process.*, vol. 184, Feb. 2023, Art. no. 109713.

[57] X. Yan, Y. Liu, and M. Jia, "Health condition identification for rolling bearing using a multi-domain indicator-based optimized stacked denoising autoencoder," *Struct. Health Monitor.*, vol. 19, no. 5, pp. 1602–1626, Sep. 2020.

[58] Y. Zhou, X. Liang, W. Zhang, L. Zhang, and X. Song, "VAE-based deep SVDD for anomaly detection," *Neurocomputing*, vol. 453, pp. 131–140, Sep. 2021.

[59] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.

[60] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Muller, and M. Kloft, "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.

[61] X. Ma, Y. Lin, Z. Nie, and H. Ma, "Structural damage identification based on unsupervised feature-extraction via variational auto-encoder," *Measurement*, vol. 160, Aug. 2020, Art. no. 107811.

[62] Z. Yuan, S. Zhu, C. Chang, X. Yuan, Q. Zhang, and W. Zhai, "An unsupervised method based on convolutional variational auto-encoder and anomaly detection algorithms for light rail squat localization," *Construct. Building Mater.*, vol. 313, Dec. 2021, Art. no. 125563.

[63] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, Aug. 1999.

[64] I. D. Khurjekar and J. B. Harley, "Closing the sim-to-real gap in guided wave damage detection with adversarial training of variational auto-encoders," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2022, pp. 3823–3827.

[65] Y. Zhang, X. Xie, H. Li, and B. Zhou, "An unsupervised tunnel damage identification method based on convolutional variational auto-encoder and wavelet packet analysis," *Sensors*, vol. 22, no. 6, p. 2412, Mar. 2022.

[66] A. N. Akansu, W. A. Serdijn, and I. W. Selesnick, "Wavelet transforms in signal processing: A review of emerging applications," *Phys. Commun.* vol. 3, no. 1, pp. 1–18, 2010.

[67] X. Yan, D. She, and Y. Xu, "Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions," *Exp. Syst. Appl.*, vol. 216, Apr. 2023, Art. no. 119479.

[68] Y. Xu, Y. Tian, and H. Li, "Unsupervised deep learning method for bridge condition assessment based on intra-and inter-class probabilistic correlations of quasi-static responses," *Struct. Health Monitor.*, vol. 22, no. 1, pp. 600–620, Jan. 2023.

[69] X. Yan, D. She, Y. Xu, and M. Jia, "Deep regularized variational autoencoder for intelligent fault diagnosis of rotor-bearing system within entire life-cycle process," *Knowl.-Based Syst.*, vol. 226, Aug. 2021, Art. no. 107142.

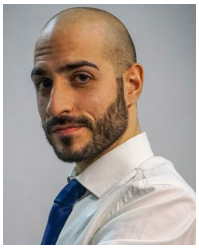
[70] D. M. Hawkins, *Identification of Outliers*, vol. 11. Berlin, Germany: Springer, 1980.

[71] C. P. Burgess, I. Higgins, A. Pal, L. Matthey, N. Watters, G. Desjardins, and A. Lerchner, "Understanding disentangling in β-VAE," 2018, *arXiv:1804.03599*.

[72] M. J. Kusner, B. Paige, and J. M. Hernandez-Lobato, "Grammar variational autoencoder," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 1945–1954.

[73] W. S. Noble, "What is a support vector machine?" *Nature Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, Dec. 2006.

[74] G. Gui, H. Pan, Z. Lin, Y. Li, and Z. Yuan, "Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection," *KSCE J. Civil Eng.*, vol. 21, no. 2, pp. 523–534, Feb. 2017.
[75] Y. Kim, J. W. Chong, K. H. Chon, and J. Kim, "Wavelet-based AR–SVM for health monitoring of smart structures," *Smart Mater. Struct.*, vol. 22, no. 1, Jan. 2013, Art. no. 015003.
[76] H. Pan, M. Azimi, G. Gui, F. Yan, and Z. Lin, "Vibration-based support vector machine for structural health monitoring," in *Proc. Int. Conf. Experim. Vibrat. Anal. Civil Eng. Struct.* Cham, Switzerland: Springer, 2017, pp. 167–178.
[77] Y. Chen, X. S. Zhou, and T. S. Huang, "One-class SVM for learning in image retrieval," in *Proc. Int. Conf. Image Process.*, 2001, pp. 34–37.
[78] S. Dyke, "Report on the building structural health monitoring problem phase 2 analytical," 2011. [Online]. Available: https://datacenterhub.org/resources/2810
[79] I. Goodfellow et al., *Deep Learning*, vol. 521. Cambridge, MA, USA: MIT Press, 2016, p. 800.
[80] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*. Berlin, Germany: Springer, 2013.
[81] J. Snoek, H. Larochelle, and R. P. Adams, "Practical Bayesian optimization of machine learning algorithms," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, 2012, pp. 1–9.
[82] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[83] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
[84] J. Tritscher, A. Krause, and A. Hotho, "Feature relevance XAI in anomaly detection: Reviewing approaches and challenges," *Frontiers Artif. Intell.*, vol. 6, Feb. 2023, Art. no. 1099521.



ANDREA POLLASTRO received the M.Sc. degree in computer science from the University of Naples Federico II, Italy, in 2019, where he is currently pursuing the Ph.D. degree with the Department of Information Technology and Electrical Engineering (DIETI). His research interests include machine learning and neural networks.

GIUSIANA TESTA is currently pursuing the Ph.D. degree with the Department of Structures for Engineering and Architecture (DIST), University of Naples Federico II, Italy. Her research interests include the health monitoring of structures through dynamic identification techniques, optimal sensors' placement methods, the application of artificial intelligence to the structural monitoring using the artificial neural networks, the safety checks of reinforced concrete existing bridges, and the estimation of mobile loads through traffic microsimulation.

ANTONIO BILOTTA is currently an Associate Professor of structural engineering. He participated and coordinated various research projects with the University of Naples Federico II aimed to developing methodologies and technologies for management and requalification of historical centers, for safety of urban systems, risk management and safety of infrastructures at regional scale, and intelligent monitoring system for urban infrastructure security. He is the author of more than 150 papers (more than 30 in peer-reviewed international journals) mainly focused on the following themes, such as bridge dynamic identification techniques in the frequency domain and machine learning techniques in the time domain, strengthening of existing reinforced concrete structures with fiber-reinforced polymers (FRP), effects of fire on concrete structures reinforced with FRP bars or strengthened with FRP systems, and effects of earthquake on structures.

ROBERTO PREVETE received the M.Sc. degree in physics and the Ph.D. degree in mathematics and computer science. He is currently an Assistant Professor of computer science with the Department of Electrical Engineering and Information Technologies (DIETI), University of Naples Federico II, Italy. His research has been published in international journals, such as *Biological Cybernetics*, *Experimental Brain Research*, *Neurocomputing*, *Neural Networks and Behavioral*, and *Brain Sciences*. His current research interests include computational models of brain mechanisms, machine learning, and artificial neural networks and their applications.

Open Access funding provided by 'Università degli Studi di Napoli "Federico II"' within the CRUI CARE Agreement