POLITECNICO DI TORINO
Repository ISTITUZIONALE

Data augmentation for damaged scenarios in floating offshore wind turbines: an approach based on diffusion architecture, hierarchical variational approximation and healthy data

Original
Data augmentation for damaged scenarios in floating offshore wind turbines: an approach based on diffusion architecture, hierarchical variational approximation and healthy data distribution / Fathnejat, H., Nava, V.. - In: ENGINEERING WITH COMPUTERS. - ISSN 0177-0667. - (2025). [10.1007/s00366-025-02148-6]

Availability:
This version is available at: 11583/2999897 since: 2025-05-06T11:32:36Z

Publisher:
Springer Science and Business Media Deutschland GmbH

Published
DOI:10.1007/s00366-025-02148-6

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

07 October 2026

Engineering with Computers
https://doi.org/10.1007/s00366-025-02148-6

ORIGINAL ARTICLE

Data augmentation for damaged scenarios in floating offshore wind turbines: an approach based on diffusion architecture, hierarchical variational approximation and healthy data distribution

Hamed Fathnejat[1] · **Vincenzo Nava**[2]

Received: 20 December 2023 / Accepted: 11 April 2025
© The Author(s) 2025

Abstract
Developing digital twin and condition monitoring models for Floating Offshore Wind Turbines (FOWTs) mooring systems requires massive data across various health, operational, and metocean conditions. The scarcity of real damage-associated data may represent a significant challenge. Deep generative models (DGMs) have recently been introduced as powerful tools for oversampling scarce data. However, most oversampling methods focus on minority intra-class information. The inter-class dynamics between minority and majority classes are often ignored, increasing the risk of overfitting, especially in scenarios with high imbalance ratios. This study proposes a novel hierarchical variational autoencoder (HVAE) utilizing the diffusion probabilistic architecture, healthy (majority) data distribution, and the relation between healthy and damage-associated data in mooring systems of FOWTs to learn the damaged state distribution. We first evaluate HVAE's ability to augment minority data based on majority distribution, using the MNIST benchmark image dataset for validation. This experiment compares the performance of HVAE with conventional and recent oversampling techniques. The second use case is the OC4-DeepCWind FOWT benchmark. The fine-tuned HVAE can generate damage-associated platform records for various sea states. Experimental results on MNIST indicate that HVAE achieves significant improvements over alternative oversampling techniques in downstream classification tasks, particularly in case of extreme imbalance. In the FOWT use case, the records generated for unseen sea states can incorporate the diversity and complexity of the majority ones, hence decreasing overfitting for the majority of sea states in downstream binary classification, highlighting the efficacy and generalization of HVAE.

Keywords Structural health monitoring (SHM) · MNIST data augmentation · Class imbalanced data · Hierarchical variational autoencoder (HVAE) · Diffusion process · Floating offshore wind

1 Introduction

Floating offshore wind turbines (FOWTs) represent the only technically viable solution for installing offshore wind turbines in deep water [1]. Until 2022, 32 FOWTs have been installed at a commercial scale worldwide [1] for a total capacity of 121 MW [2]. Nevertheless, the outlook for the next decade is that several projects are expected to be successfully deployed, reaching a capacity of 18.9 GW by 2030 [1] and 264 GW by 2050 [3]. Among the different types of floating support structures, semisubmersible platforms are among the most investigated technologies [2, 4]. Mooring systems are critically important in floating semisubmersible platforms because they provide station-keeping and operational functionality from the performance perspective [2]. However, they also considerably impact the project's economic viability [5].

Therefore, real-time condition monitoring of mooring systems is crucial for developing preventive maintenance strategies and reducing costs [6]. However, the task of building a digital twin (DT) of the mooring system is particularly challenging, as measurements of the mooring systems under different environmental and operational conditions

✉ Hamed Fathnejat
hfathnejat@bcamath.org

Vincenzo Nava
vincenzo.nava@polito.it

[1] Basque Center for Applied Mathematics, 14 Mazarredo Ave, 48009 Bilbao, Biscay, Spain

[2] Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy

[7], required for the construction of the DT [8], are not available. Indeed, records of the mooring system's behavior under different failure modes are scarce, and when available, they may be protected by confidentiality issues. On the other hand, producing records by simulating the physical model of a floating substructure is computationally expensive and accompanied by many uncertainties.

Further, when available, the monitored mooring data cannot be well-arranged and balanced concerning the different environmental and operational conditions, especially when considering damaged conditions. The so-defined class-imbalanced problem is generally challenging in the structural health monitoring (SHM) [9–18]. One of the approaches to rebalancing this imbalance is resampling the data by under-sampling the healthy state data named as the majority class or over-sampling the damaged state data named the minority class [19–21]. Recently, the new trend in the oversampling method is based on DGMs that learn the behavior and distribution of minority data and augment them by producing synthetic data. Fajardo et al. [22] investigate the performance of class conditional variants of generative adversarial networks (GANs) and variational autoencoders (VAEs) to solve the imbalance problem. The results report that the CVAEs are more robust to imbalanced datasets than their CGAN equivalents. This conclusion reverses when GANs and VAEs are utilized on the original, well-balanced MNIST and Fashion MNIST datasets. Other studies [23, 24] propose solutions for SHM based on DGMs. Soleimani-Babakamali et al. [9] addressed the challenge of an imbalanced dataset utilizing the GAN to generate the low sampled data for health state classification using long short-term memory models (LSTMs). Results demonstrate improving the classification accuracy of low-sampled classes employing the GAN-generated data. Luleci et al. [25] utilized one-dimensional Wasserstein deep convolutional GAN using gradient penalty (1-D WDCGAN-GP) to generate labeled data synthetically. After that, the augmented data generated in the previous stage were the training data set of a 1-D deep convolutional neural network for damage identification. The results showed that the 1-D WDCGAN-GP can successfully overcome damage-associated data scarcity in vibration-based damage detection applications of civil structures.

So far, VAEs are mainly employed to solve different problems belonging to the SHM domain, such as feature extraction and data space reduction [26–30]. In the context of class-imbalanced SHM problems, Zhao et al. [31] designed the normalized conditional VAE with adaptive focal loss (NCVAE-AFL) framework to make the diagnosis effect of the minority classes more desirable. NCVAE is enhanced by the batch normalized (BN) layer, which is a method of regularising the output of the encoder to prevent the disappearance of KL divergence for imbalanced training data. Besides, a new AFL has also been designed and added to NCVAE loss to reweight the imbalanced labels. Results demonstrate that the proposed algorithm significantly outperforms typical class-imbalanced diagnosis methods. Li and Betti [32] proposed a novel data augmentation framework based on conditional VAE (CVAE) architecture to generate the cepstral coefficients as response features. They defined the condition variable vector of CVAE utilizing an unsupervised learning strategy. The proposed method can augment the initial training dataset, leading to better damage classification performance than the initial training dataset.

All the studies mentioned above propose solutions limited to solving the class-imbalanced problem of DL-based damage identification/classification. Furthermore, only minority class information is applied to make DGMs learn the minority data distribution. Meanwhile, recent studies have proven that besides the minority data using majority data information, DGMs can augment the minority data with better diversity and affluence. In this regard, Qingzhong et al. [20] introduced an objective function enabling VAE to augment minority data by leveraging majority-based prior. VAEs are the DGMs generating data based on their ability to extract and construct the data's reduced feature space (latent distribution) [33].

Furthermore, an extension to the VAE is the hierarchical VAE (HVAE) concept applying more than one stochastic latent layer/distribution [34]. Besides, the diffusion process is a random process that changes the data over time. A diffusion probabilistic model (DPM) is a model that can reverse this process by learning a parameterized Markov chain. It uses a variational inference to generate samples that look like the data before the diffusion occurs [35].

The present study introduces a new hierarchical variational approximation approach based on a diffusion architecture to construct the minority data distribution utilizing both majority and minority data features. Using constructed minority data distribution, the hierarchical VAE (HVAE) generates real-scale damage-associated mooring records (minority data) under various environmental and operational conditions. Likewise, the generated responses can be employed for different downstream applications, such as solving class-imbalanced problems and building DTs. As a novelty of the proposed HVAE, based on the relation between healthy state (majority) and damaged state (minority) data, the damaged state data distribution from healthy state is obtained as a two-step reverse diffusion process, and a novel hierarchical variational approximation is proposed to parameterize this process. Implementing this concept through a two-stage pretrain-finetune training procedure facilitates the augmentation and generation of limited damaged state data from a healthy state. Specifically, the proposed approach allows data generation under various conditions to build an efficient DT of the mooring systems

of FOWTs. It allows scientists and engineers to understand better the damaged behavior of mooring systems under various intensity-varying sea states.

To validate our proposed DGM model, we employed the MNIST benchmark dataset [36], comparing our results with those provided by [20], which assesses both conventional and recent oversampling techniques for MNIST data augmentation. Furthermore, we preserved the same conditions and multiclass classification model as [20] to guarantee a fair comparison.

As the samples of real data for mooring systems are scarce, the data studied in this paper are simulated in OpenFast [37] based on the numerical model of the OC4 platform developed in the DeepCwind project [38]. Such a model represents a benchmark study validated by comparing the results with computational fluid dynamics (CFD) studies [39, 40]. Furthermore, to account for the randomness of external excitation and modeling uncertainties, the random responses produced by varying the seed value in the OpenFast simulations are considered testing data.

In this paper, the authors present a structured framework for introducing a data augmentation approach, HVAE. The paper begins with an introduction in Sect. 1, providing the necessary background and motivation for the study. In Sect. 2, the authors detail the theoretical principles and HVAE development and implementation procedure. Section 3 introduces the MNIST case study, detailing its methodology, results, and discussions. The mooring system case study follows in Sect. 4, presented in a similar structure, with its methodology, results, and discussions. The respective case studies also discuss the necessary steps for data preparation and model implementation. Finally, in Sect. 5, the outcomes are summarised, and prospective further applications of HVAE are outlined.

2 Methodology

This study introduces a novel generative model that utilizes variational autoencoder (VAE) and diffusion probability principles, along with a two-stage pretrain-finetune training framework, to reconstruct the latent distribution of minority data based on the distribution of majority data. This generative model can augment minority data by feeding corresponding majority data, as detailed in the following subsections.

2.1 Pre-training phase

Pretraining is performed on the majority class, representing the dataset's more balanced classes under various conditions or styles. The minority class demonstrates a notable imbalance and is defined by a restricted quantity of data points. The trainable weights of the standard VAE are optimized based on the Eq. (1):

$$\mathrm{ELBO}_{\theta,\phi}(X^+) = \mathbb{E}_{q_\phi}\left[\log p_\theta(X^+ \mid z_1)\right] - \mathrm{KL}\left[q_\phi(z_1 \mid X^+) \parallel p_\theta(z_1)\right] \tag{1}$$

In this step, based on Eq. (1), the conventional VAE is trained on the majority data, X^+. z_1 represents a latent variable sampled from the majority data latent/feature space, designated as the baseline latent space.

2.2 Fine-tuning phase

This section introduces the fine-tuning process to construct the latent distribution/feature space for the minority class. The fine-tuning method is based on the HVAE idea, described as a VAE with multiple stochastic latent layers and a DPM architecture. This technique assumes that the HVAE comprises two stochastic layers that generate two latent variables, z_1 and z_2. By introducing a variational approximation to the true posterior and a bottom-up formula for the inference model and by extending Eq. (1), Eq. (2) will be:

$$\begin{aligned}\mathrm{ELBO}_{\theta,\phi}(x^-) &= \mathbb{E}_{q_\phi}\left[\log p_\theta(x^-, z_2, z_1) - \log q_\phi(z_2, z_1 \mid x^-)\right] \\ &= \mathbb{E}_{q_\phi}\left[\log p_\theta(x^- \mid z_2) + \log p_\theta(z_2 \mid z_1) + \log p_\theta(z_1)\right. \\ &\quad \left. - \log q_\phi(z_2 \mid x^-) - \log q_\phi(z_1 \mid z_2)\right]\end{aligned} \tag{2}$$

In Eq. (2), x^- represents the minority data, z_1 is the latent variable corresponding to the baseline (majority) data, and z_2 denotes the latent variable associated with the minority data.

A DPM is a variant of a latent variable generative model that characterizes data through a series of $\mathcal{T}$ variables, [41]. The first variable $x_0 \sim p(x)$ represents the observed data, whereas the subsequent variables $x_{1:\mathcal{T}}$ are hidden and not directly observable.

The DPM model establishes a joint distribution $p_\theta\left(x_{0:\mathcal{T}}\right)$ that inverts a diffusion process. This procedure constitutes a Markov chain characterized by Gaussian transitions originating from $p(x_\mathcal{T}) = \mathcal{N}(x_\mathcal{T}; 0, I)$. The model acquires the ability to invert the process by optimizing a variational bound, as defined in Eq. (3):

$$-\mathrm{ELBO}_{\theta,\phi}(x_0) = \mathbb{E}_{q_\phi}\left[-\log p_\theta(x_\mathcal{T}) - \sum_{t\geq 1}^{\mathcal{T}} \log \frac{p_\theta\left(x_{t-1} \mid x_t\right)}{q_\phi\left(x_t \mid x_{t-1}\right)}\right] \tag{3}$$

For $\mathcal{T} = 2$ in Eq. (3), $x_\mathcal{T}$ is comparable to z_1 and x_0 to x^-; hence, according to Eqs. (2) and (3), the proposed loss function can be expressed as Eq. (4) and visualized in Fig. 1:

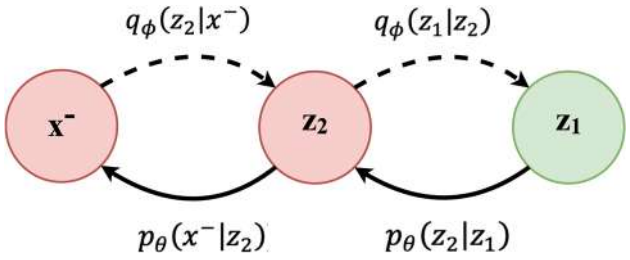

Fig. 1 HVAE loss function visualization

$$\mathrm{ELBO}_{\theta,\phi}(x^-) = \mathbb{E}_{q_\phi}\left[\log p_\theta(x^- \mid z_2)\right] - \mathrm{KL}\left[q_\phi(z_2 \mid x^-) \parallel p_\theta(z_2 \mid z_1)\right] - \mathrm{KL}[q_\phi(z_1 \mid z_2) \parallel p_\theta(z_1)] \tag{4}$$

The fine-tuning procedure presented in Eq. (4) and Fig. 1 involves the simultaneous estimation of the distributions $q_\phi(z_1 \mid z_2)$ and $q_\phi(z_2 \mid x^-)$ to obtain z_2. This procedure ensures that these distributions approximately correspond to $p_\theta(z_1)$ and $p_\theta(z_2 \mid z_1)$, respectively. Moreover, decoding z_2 should enable the reconstruction of x^-, specifically $p_\theta(x^- \mid z_2)$.

Assuming that z_1 and z_2 have the same dimensionality (D), the distribution $q_\phi(z_1 \mid z_2)$ will follow a standard normal distribution, following the assumptions used in the reparameterization of the pretrained VAE (Eq. 1). Consequently, the second KL divergence term in Eq. (4) remains unchanged throughout the fine-tuning process. Hence, based on these definitions, we proposed a novel reparameterization for fine-tuning the HVAE, as detailed in Eq. (5). Additionally, the adjustable weights of the HVAE are optimized by minimizing the loss function presented in Eq. (6).

$$z_2 = \left[\mu_1 + e^{\frac{\log(\sigma_1^2)}{2}}\right] \times \left[\mu_2 + e^{\frac{\log(\sigma_2^2)}{2}}\right] \times z_1 \tag{5}$$

In Eq. (5), μ_1 and μ_2 denote the means, while $\log(\sigma_1^2)$ and $\log(\sigma_2^2)$ indicate the log variances of the normal distributions from which the latent variables z_1 and z_2 are sampled, respectively.

$$L_{\mathrm{HVAE}}(x^-;\phi,\theta) = \mathcal{L}_{\mathrm{recon}}(x^-,\hat{x}^-) - \mathrm{KL}\left[q_\phi(z_2 \mid x^-) \parallel p_\theta(z_2 \mid z_1)\right] + \mathrm{FID}_{\mathrm{loss}} \tag{6}$$

The term $\mathcal{L}_{\mathrm{recon}}$ in Eq. (6) quantifies the difference between the input x^- and its reconstruction $\hat{x}^-$. The type of reconstruction error function in a VAE is determined based on the data's characteristics and the activation function employed in the last layer of the decoder. Binary cross-entropy (BCE) is commonly utilized for binary or binarized data, particularly with a sigmoid activation function. In contrast, mean squared error (MSE) is more suitable for continuous data, such as time-series data, typically employing tanh or linear activation functions. This decision enhances model performance by matching reconstruction loss with data characteristics and tasks.

Equation (7) defines the MSE:

$$\mathrm{MSE}(x^-,\hat{x}^-) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i^- - \hat{x}_i^-\right)^2 \tag{7}$$

In Eq. (7), x_i^- are the real input minority data, $\hat{x}^-$ are the reconstructed data, and N is the number of data points.

The binary cross-entropy (BCE) is defined by Eq. (8):

$$\mathrm{BCE}(x^-,\hat{x}^-) = -\frac{1}{N}\sum_{i=1}^{N}\left[x_i^- \log(\hat{x}_i^-) + (1 - x_i^-)\log(1 - \hat{x}_i^-)\right] \tag{8}$$

In Eq. (8), x_i^- and $\hat{x}_i^-$ represent the actual binary values and predicted probabilities, respectively, and N is the number of data points.

Furthermore, The Fréchet inception distance (FID) ([42]) loss in Eq. (6) is calculated via Eq. (9):

$$\mathrm{FID}_{\mathrm{loss}} = \|\mu_{x^-} - \mu_{\hat{x}^-}\|_2^2 + \mathrm{Tr}\left(C_{x^-} + C_{\hat{x}^-} - 2\left(C_{x^-}C_{\hat{x}^-}\right)^{1/2}\right) \tag{9}$$

In Eq. (9), μ_{x^-} and $\mu_{\hat{x}^-}$ represent the means, while C_{x^-} and $C_{\hat{x}^-}$ denote the covariance matrices of the real and generated minority data, respectively. In this context, Tr represents the trace of the matrices. The Fréchet Inception Distance (FID) serves as an index for monitoring the fine-tuning process, where reduced FID values signify more remarkable similarity between the compared datasets. The pretrain-finetune training framework of the proposed DGM, HVAE, is outlined in Algorithm 1 and depicted in Fig. 2.

Notation: The Algorithm 1 uses the following parameters:

- $X_h^+ = \{x_n^+\}_{n=1}^{N^+}$: Majority data frames, where N^+ is the total number of frames, and x_n^+ represents the n-th data frame.
- $X_d^- = \{x_n^-\}_{n=1}^{N^-}$: Minority data frames, where N^- is the total number of frames, and x_n^- represents the n-th damage frame.

- Each data frame has a shape of (Batch-size, 100, 3) for the FOWT, (Batch-size, 28×28) for MNIST. Here:
  - 100 represents time steps (from the 1st to the 100th),

- 3 corresponds to features: surge, heave, and pitch,
- 28×28 represents the pixel features of MNIST images.

3.1 Methodology and HVAE specification

The digits 0–4 were designated as the majority classes, with each class comprising 6000 samples, yielding 30,000 samples for the majority. The minority classes digits 5 through 9, were downsampled to 300 samples (60 per class) and 50 samples (10 per class), resulting in imbalance ratios of $\rho = 100$ and $\rho = 600$, respectively, consistent with [20]. The downstream classifier consists of a fully connected neural network comprising two layers, analogous to the architecture utilized in [20].

For an entirely equitable comparison, the architecture of the VAEs utilized for pre-training and fine-tuning is identical to that of the MGVAE presented in [20], which employs a multilayer perceptron (MLP)-based architecture.

Algorithm 1 HVAE Training and Testing for minority data augmentation.

Require: Majority data frames $X_h^+ = \{x_n^+\}_{n=1}^{N^+}$, Minority data frames $X_d^- = \{x_n^-\}_{n=1}^{N^-}$
Ensure: Pre-trained VAE and fine-tuned HVAE model
Initialize VAE model architecture
Set model hyperparameters (e.g., learning rate, batch size, etc.)
Split majority data frames, X_h^+, into training, validation, and test sets
Pre-training Phase:
  Train the VAE on majority data X_h^+ to learn the baseline distribution
  For each epoch:
    Update model parameters (Eq. 1)
    Track training and validation loss
  Save the pre-trained VAE model
Fine-tuning Phase:
  Assign the weights of the pre-trained VAE to the fine-tuned HVAE
  For each epoch:
    Randomly down-sample $n^+ = N^-$ frames from X_h^+
      ▷ (FOWT) Ensure that Healthy ($\{x_n^+\}_{n=1}^{n^+}$) and Damaged ($\{x_n^-\}_{n=1}^{N^-}$) data frames share the same operational conditions and time steps.
    Encode $\{x_n^+\}_{n=1}^{n^+}$ by pre-trained encoder
      Get: $\mu_h^{(n^+)} = \mu_1, \sigma_h^{(n^+)} = \sigma_1$, and $z_1^{(n^+)}$
    Encode $\{x_n^-\}_{n=1}^{N^-}$ by fine-tuning encoder
      Get: $\mu_h^{(N^-)} = \mu_2, \sigma_h^{(N^-)} = \sigma_2$
    Compute z_2 using Eq. 5
    Decode z_2 by fine-tuning decoder
      Get: $\{\hat{x_n}^-\}_{n=1}^{N^-}$
    Update model parameters (Eq. 6)
    Track training and validation loss
  Save the fine-tuned HVAE model
Validate the fine-tuned HVAE:
- Evaluate on unseen minority data utilizing the encoder of a pre-trained VAE and the encoder and decoder of a fine-tuned HVAE, and quantify performance metrics.

3 MNIST case study

This section outlines the MNIST dataset, detailing its features and the preprocessing procedures undertaken. It also describes the model implementation tailored to this dataset. Finally, the results obtained and the subsequent discussion are presented.

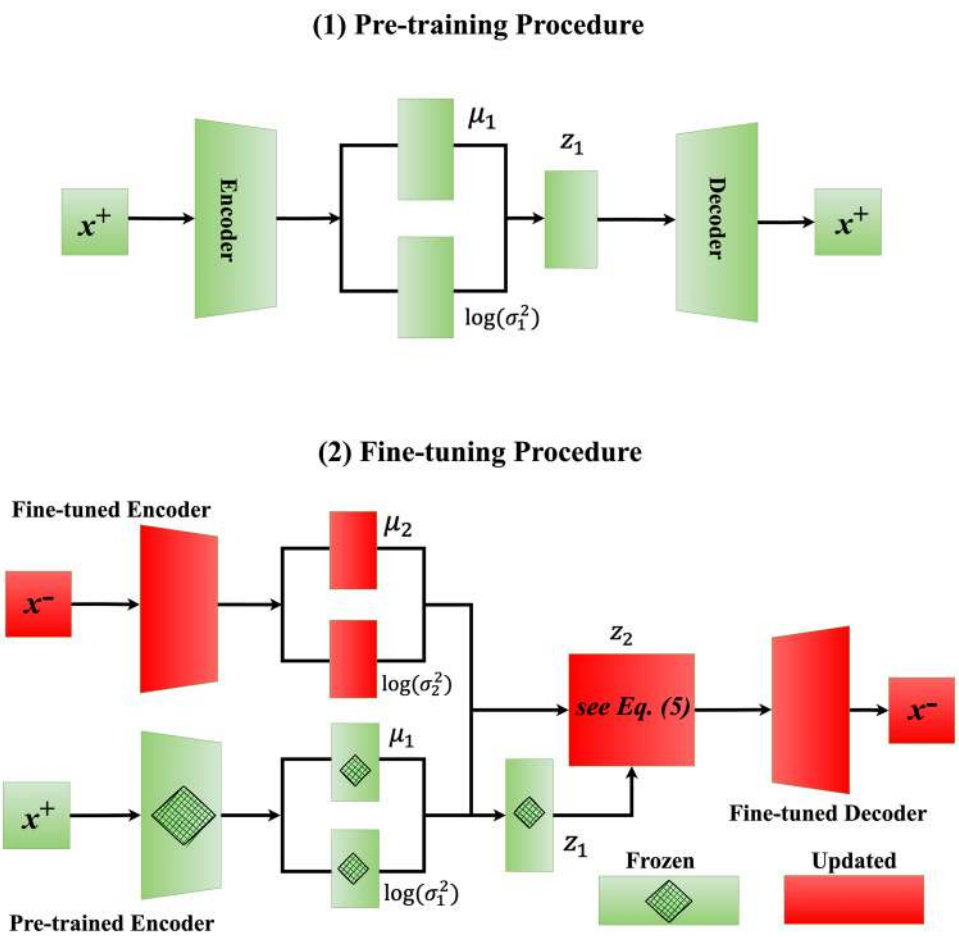

Fig. 2 HVAE training framework for minority data augmentation

The fine-tuning procedure is conducted separately for each minority class, with digits explicitly 5–9. The proposed framework for MNIST data is depicted in Fig. 3.

Figure 3 illustrates the fine-tuning procedure for the digit 5, which is categorized as a minority class. During the data generation phase, the mean value of $\mu_2 + e^{\frac{\log(\sigma_2^2)}{2}}$ is calculated. The architecture of the pre-trained and fine-tuned MLP-based VAE utilized in HVAE training is illustrated in Fig. 4.

The VAE illustrated in Fig. 4 employs Binary Cross-Entropy (BCE) as the reconstruction loss function (Eq. 8), with the Adam optimizer configured at a learning rate of 0.001. The pre-training batch size is 32 for 100 epochs. The specified batch sizes for fine-tuning each minority class are 60 ($\rho = 100$) for 10,000 epochs and 10 ($\rho = 600$) for 10,000 epochs.

3.2 Results and discussion

This section initially evaluates the effectiveness of data augmentation using HVAE through visual and statistical metrics. Secondly, we utilize the HVAE data augmentation feature for a multiclass classification task employing the MNIST benchmark image dataset [43] as a downstream application.

3.2.1 Qualitative and quantitative assessment

The qualitative results of the MNIST data augmentation are illustrated in Fig. 5.

Based on Fig. 5, HVAE generates meaningful and recognizable samples for different imbalance ratios. Both the reference majority sample and the minority sample are presented, as the generation process of HVAE is one-to-one. The upper row of each MNIST plate is the reference majority sample, while the lower row represents the generated minority. Figure 5 illustrates that the generated minority samples display stylistic similarities to the reference majority samples while possessing distinct semantic information. The writing style of MNIST numbers, characterized by tilt angle and thickness, exhibits consistency.

The statistical index, FID, derived from Eq. (9), along with MSE from Eq. (7), is provided for the generated and

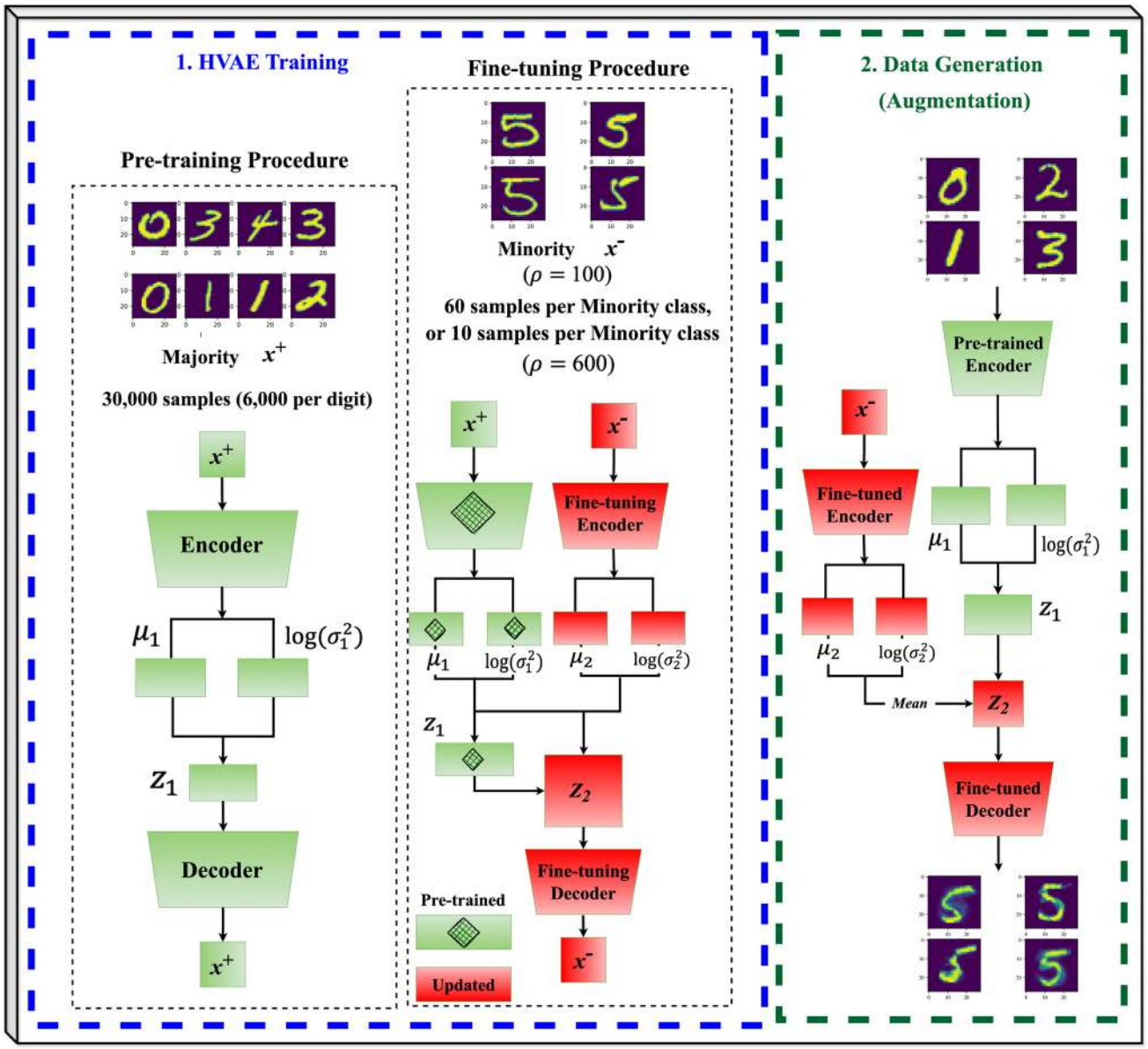

Fig. 3 The visualization of HVAE framework for MNIST data augmentation

real digits of all minority data corresponding to the majority data (30,000 samples) in Table 1.

Based on Table 1, MSE has lower sensitivity when comparing real and generated image data. FID better reflects the quality of generated image data concerning the imbalance ratio.

3.2.2 Downstream application study (behavioural assessment)

We utilize three evaluation metrics to assess the classification performance of all methods on a balanced test set, as outlined in [20]. Balanced Accuracy (B-ACC) [44, 45], Average Class Specific Accuracy (ACSA) [44, 46], and Geometric Mean (GM) [47] are metrics used for evaluating classification performance. B-ACC is the standard accuracy metric for balanced datasets; however, ACSA and GM are class-agnostic metrics, making them more appropriate for evaluating performance in imbalanced data contexts. The metrics are calculated using previously unexamined real test data, with the average results from three separate classification trials shown in Table 2 and illustrated in Fig. 6.

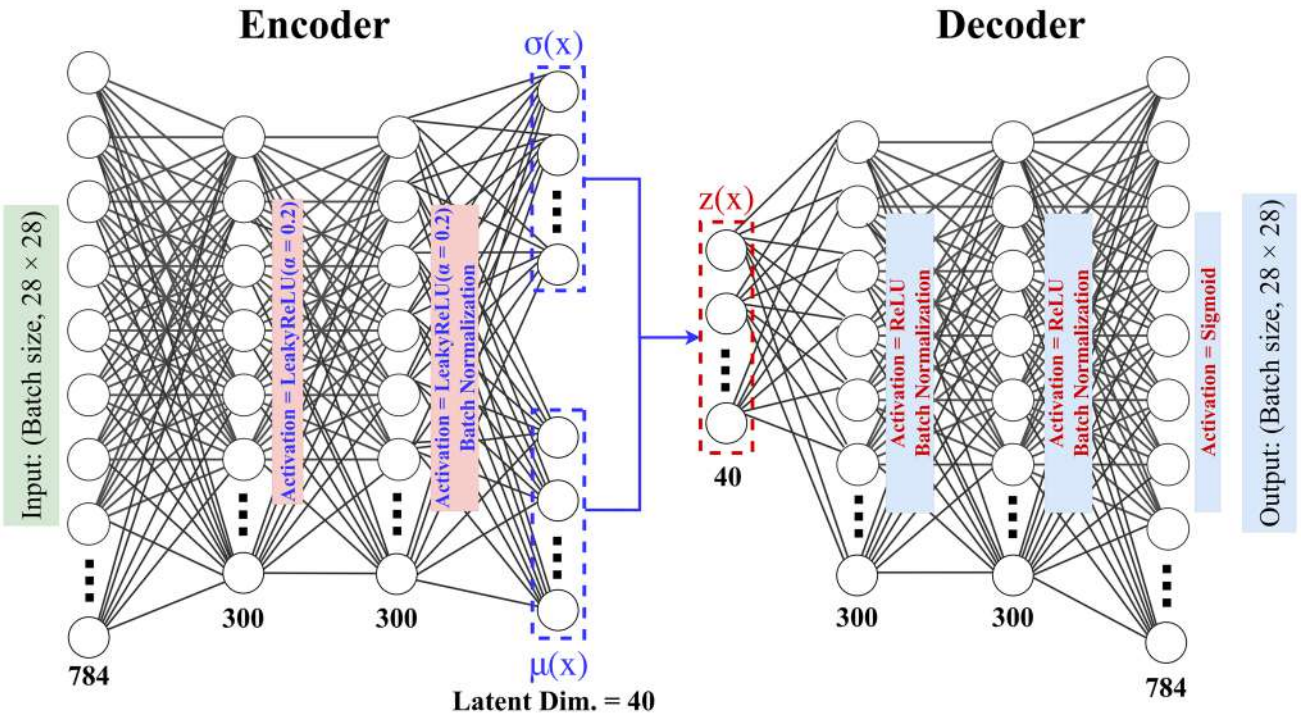

Fig. 4 The architecture of the MLP-based VAE employed in HVAE training

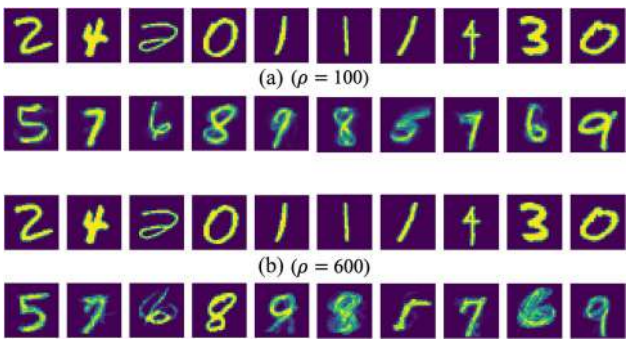

Fig. 5 Visualization of MNIST data generation by HVAE; **a** $\rho = 100$; **b** $\rho = 600$

Table 1 Quality quantification of minority MNIST data generation using HVAE

Imbalance ratio	MSE	FID
$\rho = 100$	0.0794	0.1292
$\rho = 600$	0.0852	0.1506

A concise overview of the baseline oversampling methods, utilized for comparison with the performance of HVAE in Table 2, is presented below:

1. Empirical risk minimization (ERM) is the conventional training approach utilizing cross-entropy loss without balancing procedures.
2. Random over-sampling (RS) [48]: a method for balancing the sampling distribution through repeated random sampling of the minority class.
3. SMOTE [21]: balancing the sampling distribution through linear interpolating nearest neighbors within the minority class.
4. Re-weighting (RW) [44]: adjusting the objective function based on the size of class samples.
5. Class-balanced re-weighting (CBRW) [49]: is a modified version of re-weighting (RW), which introduces the effective sample number $E_k = \frac{1-\beta^{N_k}}{1-\beta}$ for each class, with β set to 0.9999.
6. FOCAL [50]: seeks to balance the sample-wise classification loss during model training by reducing the weight of well-classified samples.
7. LDAM [51]: a margin loss aware of label distribution, promoting larger margins for few-shot classes.
8. OCVAE [22]: This method includes over-sampling the minority class using a Conditional VAE, which is a generative model that augments the dataset conditioned on class labels.
9. OCGAN or OCDCGAN [22]: This method includes oversampling the minority class using a Conditional

Table 2 Comparison of classification performance on MNIST: results of a classifier trained on balanced data augmented using the HVAE vs. conventional data augmentation techniques (two imbalance ratios)

IR	$\rho = 100$			$\rho = 600$		
Methods	B-ACC	ASCA	GM	B-ACC	ASCA	GM
ERM	51.4 ± 0.0	50.0 ± 0.0	0.0 ± 0.0	51.4 ± 0.0	50.0 ± 0.0	0.0 ± 0.0
FOCAL	51.4 ± 0.0	50.0 ± 0.0	0.0 ± 0.0	51.4 ± 0.0	50.0 ± 0.0	0.0 ± 0.0
RW	77.4 ± 1.2	76.7 ± 1.2	73.1 ± 1.3	59.9 ± 1.6	58.8 ± 1.4	41.7 ± 1.8
CBRW	75.1 ± 0.8	74.3 ± 0.7	69.8 ± 1.3	56.1 ± 0.5	55.1 ± 0.3	31.2 ± 1.2
LDAM	82.9 ± 0.5	82.4 ± 0.6	80.3 ± 0.7	63.1 ± 0.9	62.0 ± 0.8	48.7 ± 1.0
RS	79.2 ± 0.3	78.5 ± 0.3	75.4 ± 0.2	58.5 ± 1.0	57.3 ± 1.1	37.7 ± 2.0
SMOTE	80.6 ± 0.3	80.1 ± 0.2	77.4 ± 0.1	60.0 ± 0.9	58.3 ± 1.0	40.2 ± 1.1
OCVAE	83.0 ± 0.4	82.6 ± 0.4	80.6 ± 0.6	63.8 ± 0.2	62.8 ± 0.5	50.7 ± 0.5
MGVAE	85.0 ± 0.2	84.6 ± 0.2	83.2 ± 0.2	65.4 ± 1.0	64.4 ± 1.1	53.4 ± 1.1
HVAE	$\mathbf{87.0 \pm 0.2}$	$\mathbf{86.5 \pm 0.2}$	$\mathbf{85.6 \pm 0.2}$	$\mathbf{71.9 \pm 0.3}$	$\mathbf{71.0 \pm 0.4}$	$\mathbf{63.6 \pm 1.0}$

Bold values denote the best-performing results, which are achieved by HVAE

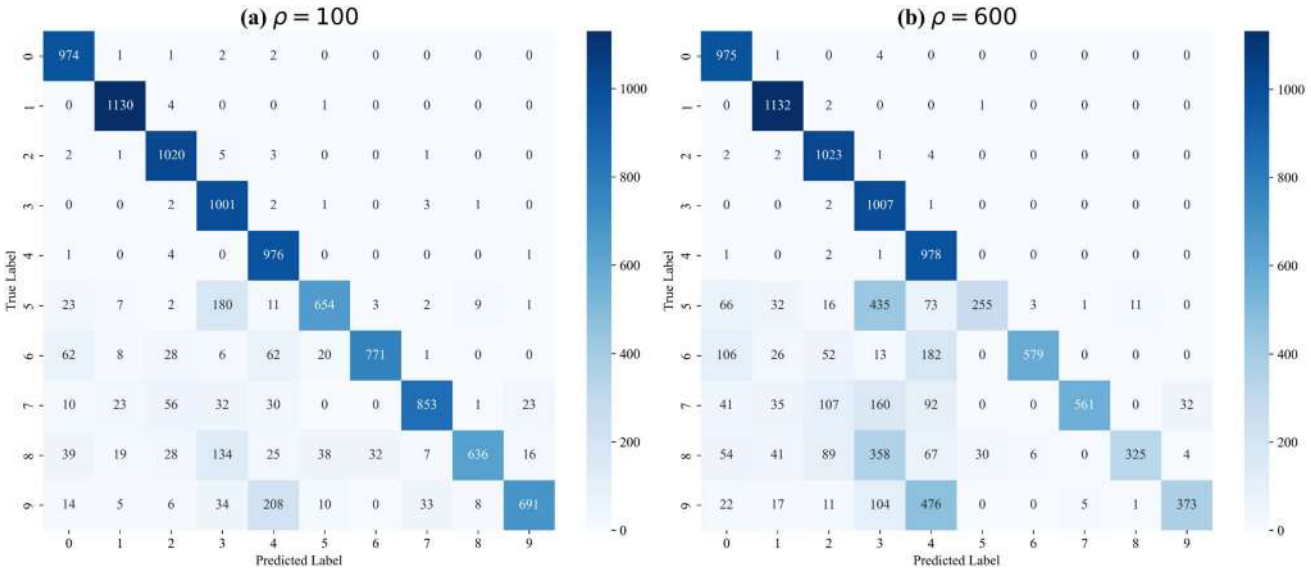

Fig. 6 Confusion matrix: multiclass MNIST classification with balanced data using the HVAE for data augmentation; **a** IR = 100, **b** IR = 600

GAN, which is a generative model that augments the dataset based on class labels.

10. MGVAE [20]: The Majority-Guided VAE presents an objective function with Elastic Weight Consolidation (EWC) regularization that facilitates the augmentation of minority data through the utilization of a majority-based prior.

Table 2 indicates that our methodology outperforms the previously mentioned techniques, yielding superior results across all comparisons. OCGAN's results were excluded due to its ineffectiveness in training on small-scale data. In instances of significant imbalance ($\rho = 600$), the GM term exhibits a percentage improvement of 10.2% compared to MGVAE. This improvement demonstrates that HVAE exhibits significantly greater effectiveness in scenarios of extreme imbalance.

The monitoring indicators for HVAE fine-tuning, calculated using Eqs. (6), (8), and (9), are presented in Fig. 7 (for $\rho = 100$) throughout the training process. The reconstruction and FID losses approach zero, as shown in Figs. 7a and b. The optimized HVAE improves the structure of minority data by iteratively refining information during each epoch, resulting in decreased losses. Figure 7c illustrates the convergence of the KL loss to a minimal, non-negative value, signifying the effective learning of the latent variable z_2 (Eq. 5). Establishing z_2 during fine-tuning serves as a regularizer to mitigate overfitting in the decoder of the fine-tuned HVAE, facilitating the generation of diverse samples of minority data despite the constraints posed by the limited examples utilized in the fine-tuning process. Diversity is illustrated in Fig. 5 and in the findings shown in Table 2.

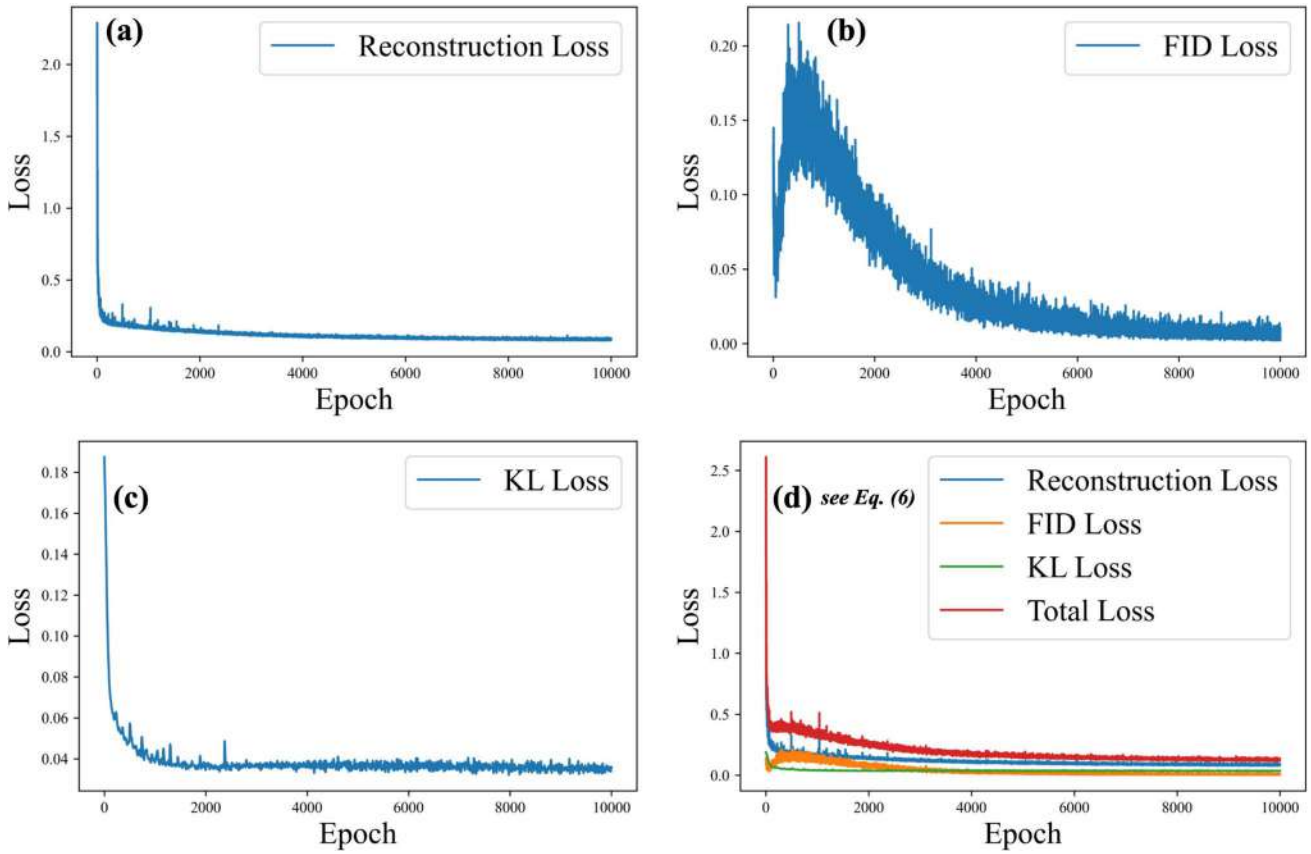

Fig. 7 Fine-tuning monitoring plots for MNIST data using the HVAE (IR = 100)

4 FOWT case study

This section outlines the characteristics and specifications of the numerical model for the DeepCwind OC4 platform, as shown in Fig. 8. It details the necessary preliminary steps for data preparation and the implementation of the HVAE model. The findings are then presented, followed by an in-depth discussion.

4.1 Methodology and HVAE specification

The motions of the semi-submersible floating system, as illustrated in Fig. 8, are regarded as variables for assessing the health status of the mooring system. The turbine's floater dynamic response u is derived from the functional outlined in Eq. (10):

$$u = F(P, r) \tag{10}$$

where in Eq. (10), P is the structural integrity of the platform's mooring lines, including structural properties, characteristics, and behavior; r represents the environmental and operational conditions affecting the floater's behavior. F is a simulation functional that could simulate the behavior of the floater based on the OpenFast tool.

The platform has six degrees of freedom (DOFs) categorized as surge, sway, and heave for translational motions and roll, pitch, and yaw for rotational motions, as illustrated in Fig. 8. This investigation examines the surge, heave, and pitch degrees of freedom (DOFs) as they are excited DOFs for three reasons:

- Waves, current, and wind are aligned in the x-direction.
- The waves are characterized as cylindrical, meaning they possess an infinite wavefront.
- The mooring/substructure/turbine system exhibits symmetry concerning the x-axis.

The time-series responses as the training, validating, and testing data are obtained under different randomly sampled combinations of the environmental and operational conditions, including wave-significant height (H_s) ranging from 1 to 7 m, peak period (T_p) between 8 and 15 s, wind speed (V) from 1 to 10 m/s, and current (C) from 0.5 to 1.5 m/s, which altogether account for the variability of the excitation. The simulation generates a response time history of over 1800s, utilizing a constant sampling frequency of 40 Hz.

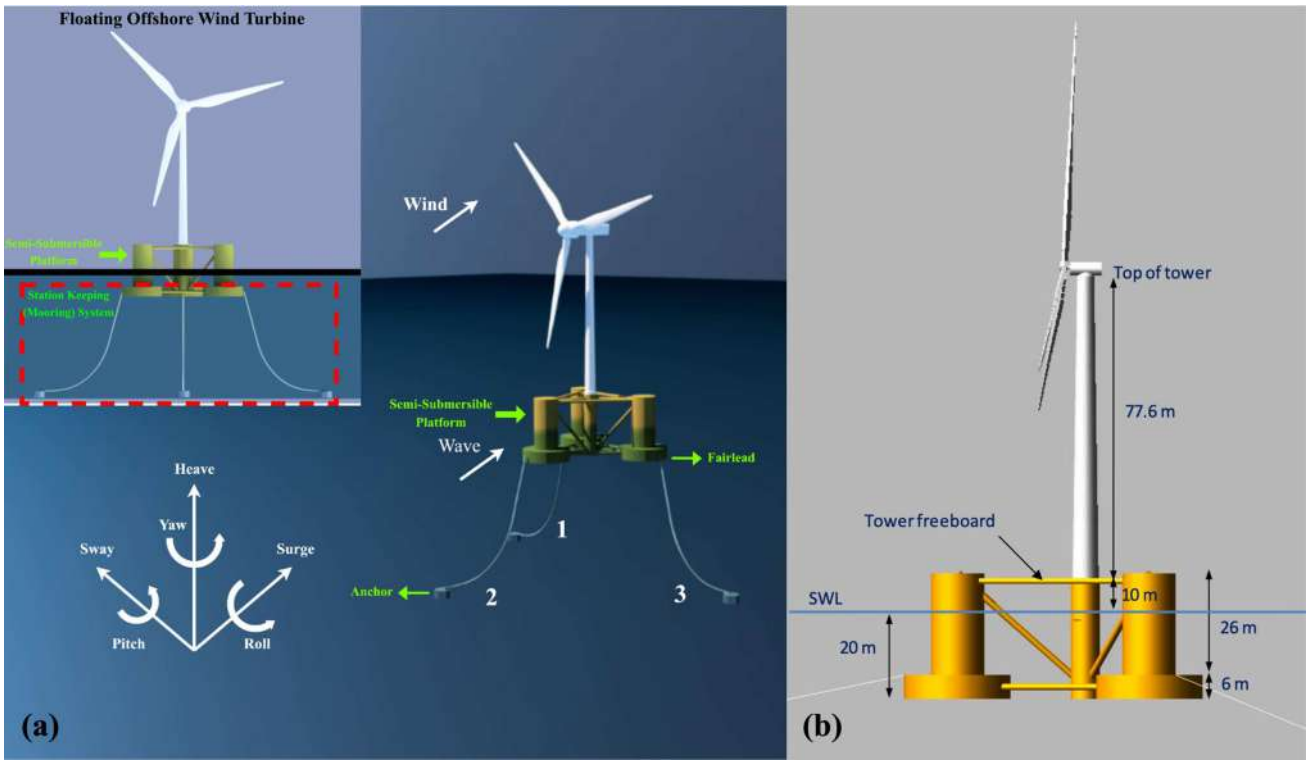

Fig. 8 Schematic illustration (**a**) and geometric dimensions (**b**) of the OC4-DeepCWind semisubmersible model [38]

The sampled response is downsampled to 5 Hz to create a realistic scenario, with the initial 300 s excluded to mitigate turbulence effects.

This study considers two types of damaged mooring lines as in [52]: biofouling and anchorage. Biofouling is the growth of mussels, algae, and other marine organisms on the mooring systems, and it has been modeled as an increase of mooring lines' mass per unit length and diameter. Anchorage status refers to the movement or shift of a line's anchor in any direction, assuming a flat seabed. The mooring lines' properties and coordinates for healthy status are presented in Table 3.

Hence, each data sample consists of a 1500-second time series of a floater's response, comprising 7500 time steps (5 Hz) and three features: surge, heave, and pitch degrees of freedom for each sea state. The VAE input layer processes 100 time steps, established through trial and error, resulting in an input shape of (batch size, 100, 3, 1), as detailed in Sect. 2.2, with parameters utilized in Algorithm 1. Tabular data is input into a two-dimensional convolutional neural network (2DCNN) utilizing a single channel. All data frames undergo shuffling before the training process. No normalization is applied during pre-training and fine-tuning to maintain real-scale outputs and improve latent space adaptation to input data.

The dataset consists of healthy samples (majority data) under 625 environmental and sea conditions ($N^+ = 625$) and differing amounts of damaged samples (minority data) ($N^- = 3, 6, 12$), which correspond to 0.5%, 1%, and 2% of the healthy state data. To model biofouling damage, the

Table 3 Properties and coordinates of mooring lines; a baseline (healthy) status

Properties of mooring lines			Coordinates of mooring lines			
Property	Value	Unit	Point	X (m)	Y (m)	Z (m)
Number of lines	3	–	Line 1 Anchor	418.80	725.38	−200.00
Segments per line	20	–	Line 1 Fairlead	20.43	35.39	−14.00
Line diameter	7.66×10^{-2}	m	Line 2 Anchor	−837.60	0.00	−200.00
Mass/length	1.13×10^{2}	kg/m	Line 2 Fairlead	−40.87	0.00	−14.00
Axial stiffness (EA)	7.54×10^{8}	N	Line 3 Anchor	418.80	−725.38	−200.00
Unstretched length	8.35×10^{2}	m	Line 3 Fairlead	20.43	−35.39	−14.00

MoorDyn module's mass per unit length for all three mooring lines is increased by 3% and 5%. Anchorage damage refers to the movement or shift of a line's anchor in any direction by ±3 m, assuming a flat seabed.

In the pre-training phase, 80% of the healthy state data (N^+) is allocated for training and validation, with 80% designated for training and 20% for validation. The remaining 20% is set aside for testing. The architecture utilized for HVAE training features a pre-trained and fine-tuned CNN-based VAE, as illustrated in Fig. 9.

Figure 10 depicts the fine-tuning process for a damage scenario within the mooring system of the OC4 platform, classified as a minority class. During the data generation phase, the mean value of $\mu_2 + e^{\frac{\log(\sigma_2^2)}{2}}$ is calculated.

The parameters for the two-dimensional convolutional neural network (2DCNN) utilized in the model in Fig. 9, are outlined as follows: - Convolution window/kernel dimensions: 2×2 - Convolution stride: 1×1 - Padding: same. The convolution stride size for 2DCNN in the second layer of the encoder and the 2DCNNTranspose in the layer preceding the output layer of the decoder is 2×1. The reconstruction loss function used is Mean Squared Error (MSE) (Eq. 8), and the optimizer applied is Adam, set with a learning rate of 0.001. The pre-training batch size is set to 32 for 100 epochs. The batch size for fine-tuning is determined based on the size of the minority data. For example, when $N^- = 6$, the total number of time steps is 6×7500, resulting in a batch size of $(6 \times 75, 100, 3, 1)$ for 1000 epochs.

The inherent three-dimensional structure of the data makes it suitable for processing with a 2DCNN. After evaluating MLP-based and 1DCNN-based architectures, the 2DCNN demonstrated superior performance. The ability to capture spatial patterns across both time and feature dimensions enhanced learning outcomes and improved the accuracy of results. By examining various configurations for this dataset, it was established that employing 100 time steps per frame alongside a 2×2 kernel size produced the most favorable outcomes.

The selection of the kernel size involved the evaluation of several alternatives, specifically 3×3, 2×1, and 3×1, given that the second dimension in each frame is 3. After extensive experimentation, it was concluded that the 2×2 kernel size produced the best metrics, such as the Fréchet

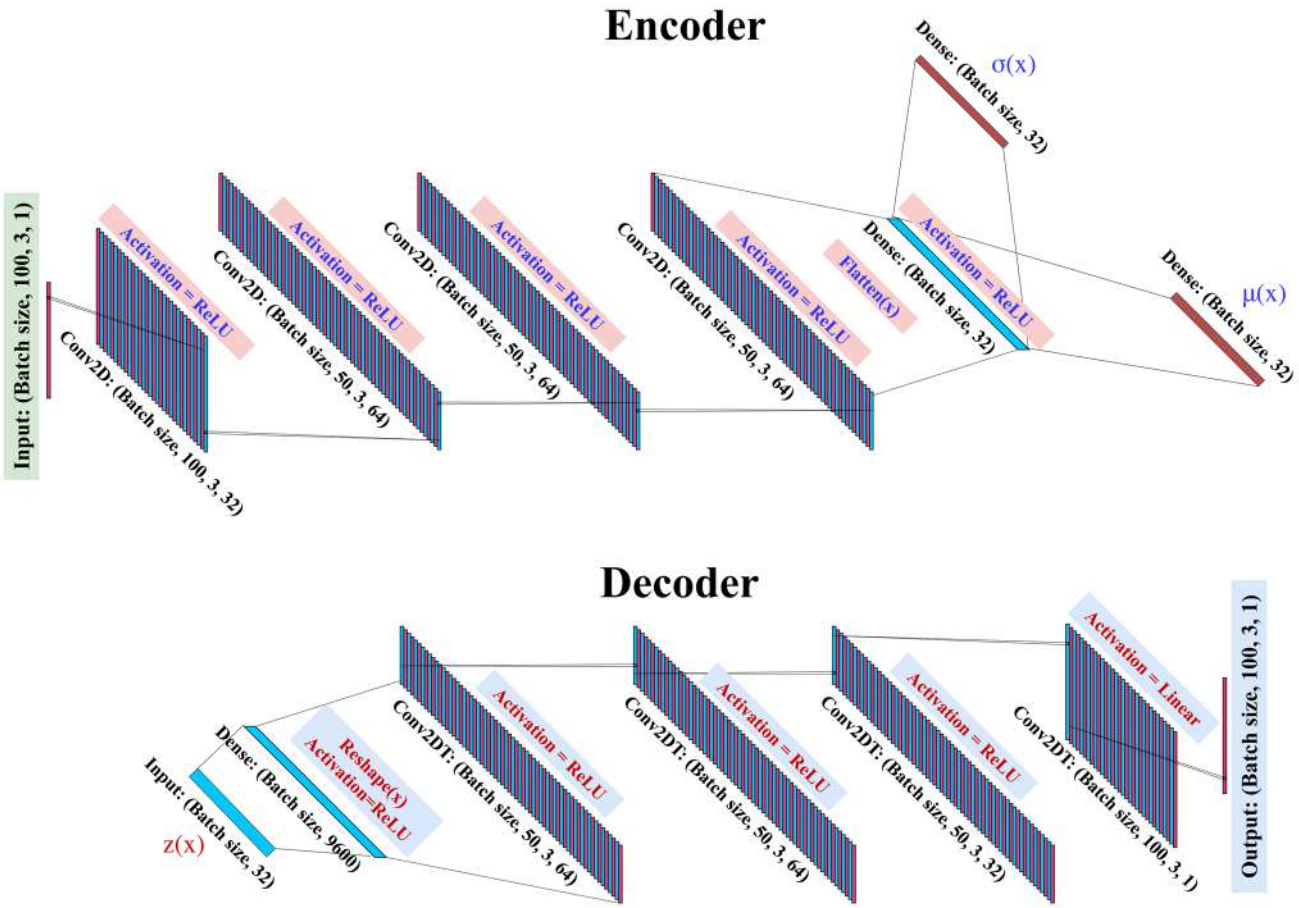

Fig. 9 The architecture of the CNN-based VAE employed in HVAE training

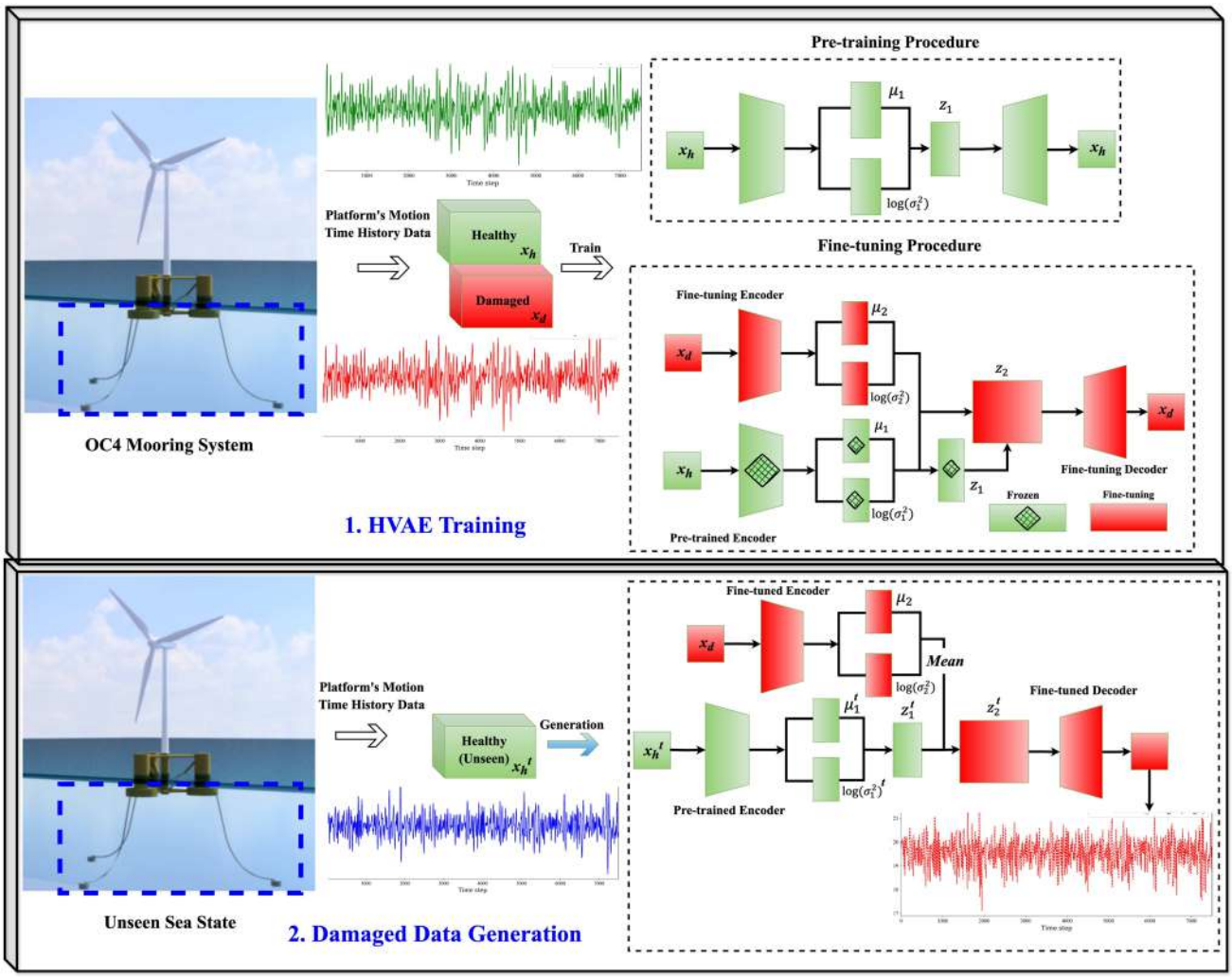

Fig. 10 The visualization of HVAE framework for FOWT mooring system data augmentation

Inception Distance (FID) and reconstruction error, particularly when generating damage-associated time series under unforeseen operational conditions.

Furthermore, the analysis of metric values indicated that utilizing four convolutional layers resulted in better results than alternative configurations. Regarding the activation function, we noted that the ReLU activation function was more effective relative to LeakyReLU inside our particular CNN design. Depending on the data type and model architecture, LeakyReLU may produce better results in some contexts.

All deep learning models were trained using Apple M1 Pro with 10 CPU cores. The memory capacity is 16 GB. The models are formed on the platform of TensorFlow version

Table 4 Performance of HVAE for three minority data sizes (biofouling 3%): mean and MAD of reconstruction error across six samples

Minority data size	Unseen response (different seed)		All 625 responses		Fine-tuning time (s)
	Mean	MAD[a]	Mean	MAD	
$N^- = 3$ (0.5%)	0.0065	0.00192	0.0097	0.0130	168
$N^- = 6$ (1%)	**0.0048**	**0.0002**	**0.0010**	**0.0001**	**287**
$N^- = 12$ (2%)	0.0046	0.0001	0.0008	0.0001	531

[a]Mean absolute deviation

Table 5 Results of generating unseen responses using HVAE for biofouling (3% and 5%) and anchorage

Damage status	Same seed		Different seed		Fine-tuning time (s)
	MSE	FID	MSE	FID	
Biofouling 3%	0.0010	0.0005	0.0044	0.0013	287
Biofouling 5%	0.0014	0.0006	0.0050	0.0020	289
Anchorage	0.0032	0.0019	0.0054	0.0025	286

2.11.0 using Keras application programming interface (API) 2.11.0. Furthermore, the OpenFAST simulations were conducted on the DIPC cluster [53].

4.2 Results and discussion

This section assesses the efficacy of data augmentation via HVAE utilizing visual and statistical metrics. Moreover, we develop a real-time monitoring model incorporating domain adaptation to accommodate the varied operational conditions of floating offshore wind turbines' mooring systems.

4.2.1 Qualitative and quantitative assessment

This section evaluates the response of a FOWT platform under an unforeseen operational and environmental combination ($V = 3.25$, $H_s = 4.0$, $T_p = 13.25$, $C = 0.5$) to assess data generation performance both visually and statistically. Additionally, to examine the impact of the randomness of external excitation by altering the wave seed in the HydroDyn module in OpenFast, the response of the healthy state corresponding to the previously mentioned environmental combination across different seed values is input into a fine-tuned HVAE to assess its accuracy in generating the corresponding damage state response.

A sensitivity analysis was conducted to determine the optimal minority data size that yields more stable results. The mean absolute deviations (MADs) and the mean of six reconstruction errors, represented as a mean squared error (MSE), for minority data sizes $N^- = 3$, $N^- = 6$, and $N^- = 12$ are computed for a biofouling severity of 3% and presented in Table 4.

Table 4 presents a thorough analysis to determine the optimal value of N^-. The results indicate that the MAD and the mean values of reconstruction errors decrease as the N^- value increases. A lower MAD indicates improved stability of generation accuracy in response to operational and environmental variations, ultimately leading to enhanced generalization of HVAE. Considering the results, the appropriate mean and MAD values for generating an unseen response, as well as for the 625 damaged state responses and the fine-tuning duration, indicate that $N^- = 6$ is a suitable choice. Although lower errors are achieved for N^-=12, the bolded values reflect the optimal trade-off between accuracy and fine-tuning cost. Table 5 presents the results of varying biofouling severity and anchorage statuses, which are fine-tuned by $N^- = 6$.

According to Table 5, biofouling with 5% severity is also tested to evaluate HVAE performance in intense damage severities. It is rare for all three mooring lines to increase mass density by more than 5 kg/m. However, this study evaluates HVAE performance for generating damage data with this extensive damage severity. Despite severe damage, the HVAE can achieve acceptable accuracy employing six damaged state responses like biofouling 3%. Additionally, HVAE fine-tuned by six damage-associated responses generates all 625 biofouling 5% state responses with total reconstruction error = 0.00162 and FID = 0.00073 differences from real ones. In the case of anchorage, the HVAE performs well for different seed-produced responses, employing six damaged state responses. Additionally, HVAE fine-tunes all 625 anchorage state responses with total reconstruction error = 0.00134 and FID = 0.00124. The generated and real responses for biofouling 5% and anchorage statuses are shown in Fig. 11 to demonstrate HVAE's efficiency with only six damaged state responses.

4.2.2 Downstream application study (behavioural assessment)

This section examines data-driven binary damage identification as a downstream application. We train the classifier on a balanced dataset obtained by augmenting generated data across two distinct imbalance ratio scenarios. The classification efficacy of the developed model is assessed using real testing data, analogous to the approach employed in Sect. 3.2.2 with MNIST data. Randomization's impact is evaluated by analyzing testing responses from diverse seed values. The proposed DGM, HVAE, can be a surrogate model utilizing Artificial Neural Networks (ANNs) to generate real-scale damage-related data. This application utilizes an ANN surrogate model alongside meta-heuristic algorithms to estimate relevant sea state parameters, analogous to the approach used for damage severity estimation [15, 17, 18, 54].

The architecture of the classifier, which is a fully connected network, is: C1 = MLP [Input dim−256], C2 = MLP [256−128], C3 = MLP [128−1]. The activation function for C1 and C2 is ReLU, while the final dense layer employs a sigmoid activation function. The binary Cross-Entropy loss function is employed with Adam optimizer and learning rate of 1×10^{-3}. The output of the sigmoid function is defined here as a prediction score between 0 (healthy) and 1 (damaged).

The classification model uses healthy and biofouling/anchorage-damaged statuses from 625 environmental and operational conditions. Eighty percent of the responses are training and validation, while the remaining twenty percent are testing. For biofouling, the input dimension is set as (batch size, frame size × num features) = (batch size, 250×3). The model is trained over 20 epochs with a batch size of 150. For anchorage, a frame size of 1250 and a batch size of 200 are used, and the classifier is trained for 25 epochs. Each feature is normalized

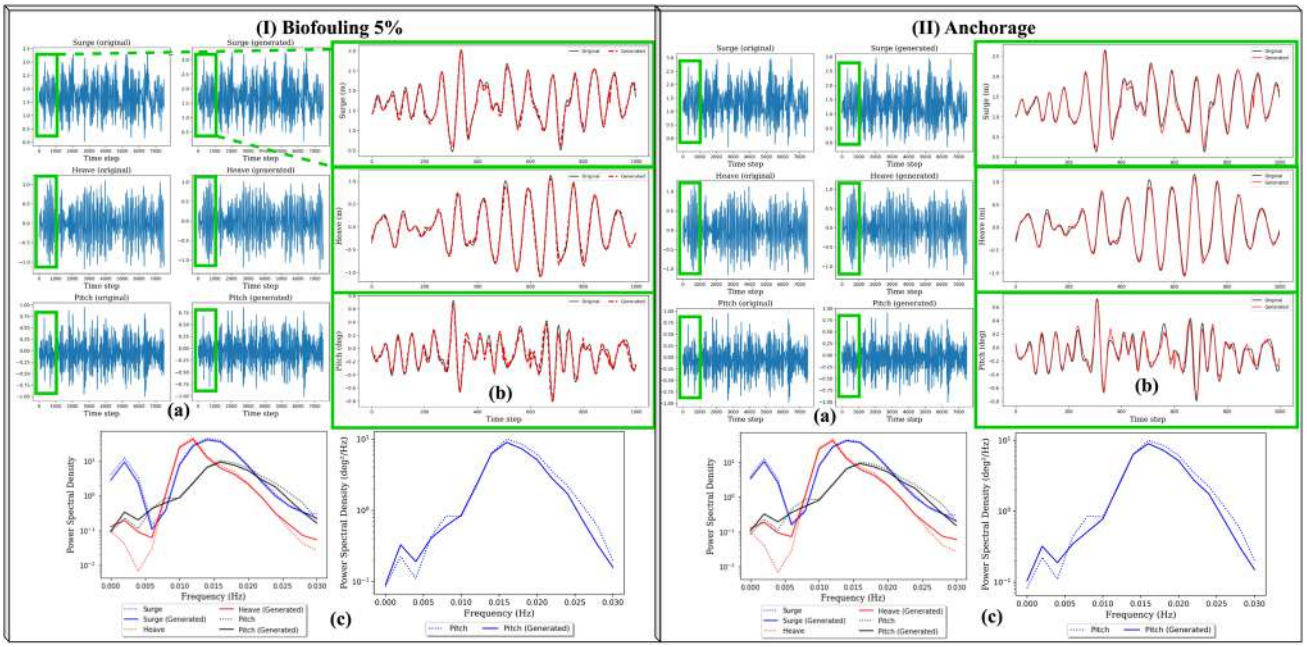

Fig. 11 Generated (I) Biofouling (5%), (II) Anchorage response using HVAE model with different seed values; **a** Unseen damaged state data sample. **b** The first 1000 time steps of unseen damaged state data sample. **c** Power spectral density

to a bounded range to enhance classification performance. Robust Scaling proves effective in this classification task, particularly in the presence of outliers. This method subtracts the median value from the feature data and divides the result by the data's interquartile range (IQR).

Thirty-one responses, each corresponding to distinct sea states with varying seed values, are randomly selected as the testing dataset. To accurately comprehend the classification results, the values of sea states that correspond to damage-associated responses, totaling 22, and the sea state values corresponding to $N^- = 6$ responses utilized for HVAE fine-tuning are presented in Tables 6 and 7, respectively.

Figure 12 illustrates the prediction scores for the binary classification of healthy (0) and damaged (1) responses in a scenario of biofouling. The classifier was trained on three datasets with differing ratios of real and synthetic data:

1. 1% Real Data + 99% Generated Data: The dataset comprises 1% real damaged data and 99% generated damaged data using HVAE (fine-tuned on 1% real damaged data).
2. 10% Real Data + 90% Generated Data: The dataset comprises 10% real damaged data and 90% generated damaged data using HVAE fine-tuned on 10% real data.
3. 100% Real, Balanced Data: This dataset includes only real damaged data and is balanced to assess classifier performance under ideal conditions.

The results are based on damaged test datasets corresponding to sea states of Table 6. An increase in real samples improves the classification model's generalization ability, leading to a consistent enhancement in prediction scores compared to the 1% real data instance.

Figure 13 presents the prediction scores for anchorage. In this case, randomization significantly influences the effectiveness of the classification algorithm, even when employing solely real data. When utilizing 1% of real data, the resulting prediction scores are generally inferior to those obtained from larger proportions of real data. An analysis of the sea states in the fine-tuning dataset (1% real data, Table 7) compared to those in the testing dataset (Table 6) indicates that the classifier exhibits reduced accuracy for sea states 8 to 10, as well as 14 and 15, as shown in Fig. 13. Sea states 10, 14, and 15 demonstrate more significant variability in wind velocity (V). The difference between these states and those employed for fine-tuning is more significant. Consequently, fine-tuning the HVAE with merely 1% of the actual damaged data causes its performance to be more vulnerable to variability than fine-tuning with a larger, more representative dataset.

The comparison of biofouling (Fig. 12) with the anchorage damage scenario (Fig. 13) reveals that anchorage damage causes a more complex classification challenge, as substantial uncertainty reduces the classifier's prediction scores, even when utilizing only real data.

Utilizing 10% real data enhances the generality of the HVAE in damaged data augmentation, particularly regarding

Table 6 Environmental and operational conditions for the damage-associated testing dataset

Sea state	Sea state combination number																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
V	10	1	10	3.25	10	7.75	7.75	10	7.75	1	10	10	10	10	10	7.75	3.25	7.75	7.75	7.75	1	5.5
H_s	7	7	1	5.5	4	4	5.5	2.5	1	2.5	4	5.5	2.5	1	1	5.5	1	1	5.5	2.5	1	1
T_p	13.25	15	9.75	13.25	9.75	11.5	13.25	8	11.5	9.75	9.75	11.5	11.5	8	9.75	13.25	9.75	13.25	11.5	11.5	15	13.25
C	1.25	1	0.75	1.5	0.75	0.75	0.5	1.25	1.5	1.5	1	1.5	1	0.75	0.75	0.75	1	1	0.75	0.75	0.5	0.75

sea states 8 to 10, as well as 14 and 15. Increasing the percentage of real data used for training enhances the HVAE's capacity to produce more precise and generalized damaged data, especially for sea states characterized by more significant fluctuation.

Furthermore, both biofouling and anchorage damage statuses indicate that more critical sea states are observed when environmental and operational conditions (sea states) feature more severe wave conditions (H_s and T_p) combined with high wind velocity (V) or high current velocity (C). Under such conditions, classification and data generation accuracy decrease, which is more pronounced when accounting for the inherent randomness in the wave.

Table 8 presents the binary classification metrics for two damage scenarios: Biofouling and Anchorage. The table evaluates the classifier's efficacy across three training datasets: 1% real, 10% real, and 100% real data. The evaluated measures are Accuracy (ACC), Precision (PREC), Recall, and Area Under the Curve (AUC).

The findings presented in Table 8 indicate that while the HVAE contributes to augmenting limited real-world data, the addition of more significant amounts of real data to the training set significantly enhances the classifier's generalization ability, particularly in challenging scenarios such as Anchorage.

4.3 Ablation study

This section conducts component analysis to examine the significance of each key component of the proposed method. Table 9 presents an investigation into the role of sampling z_1 rather than a standard normal distribution (ϵ) in Eq. 5. The results indicate that substituting the prior with a standard normal distribution instead of a healthy data distribution, sampled by variable z_1, significantly reduces the data generation performance for damaged states. This observation confirms the efficacy of the healthy state data distribution as a surrogate and initial reference for exploring the space of HVAE trainable weights. Subsequently, the standard VAE is trained on six damaged state responses related to anchorage without utilizing a pre-trained model. An unseen damaged state response, generated with a different seed value, is input into this model for generation attempts. The findings are presented in the second row of Table 9. Although the damage response is input into the VAE, accurate reconstruction is not achieved. Consequently, the accuracy of the VAE is lower than that of the HVAE, which produces this response from a healthy state response. The generalization of VAE is significantly reduced. The results confirm the efficacy of the pretrain-finetune procedure, particularly in scenarios where the minority data size is limited.

Table 7 Environmental and operational conditions of the damage-associated HVAE fine-tuning dataset

Sea state	Sea state combination number					
	1	2	3	4	5	6
V	7.75	7.75	5.5	5.5	1	1
H_s	5.5	7	2.5	1	2.5	4
T_p	15	8	13.25	13.25	11.5	15
C	1.5	1.5	0.5	1.5	1.25	1

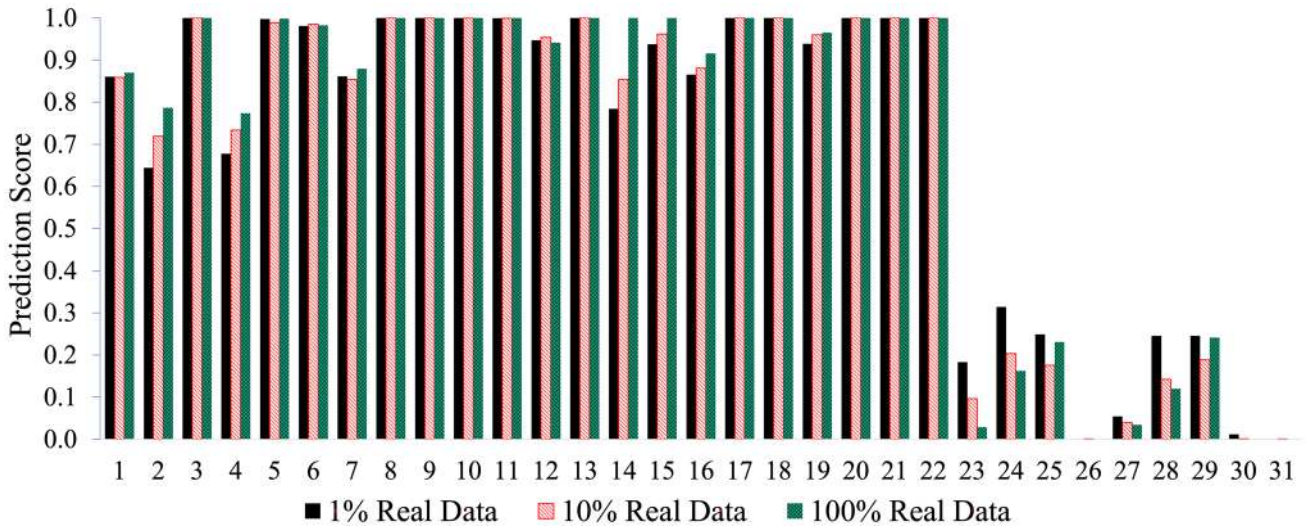

Fig. 12 Biofouling; Prediction scores for binary classification of healthy (0) and damaged (1) responses, utilizing real and HVAE-generated data at different ratios

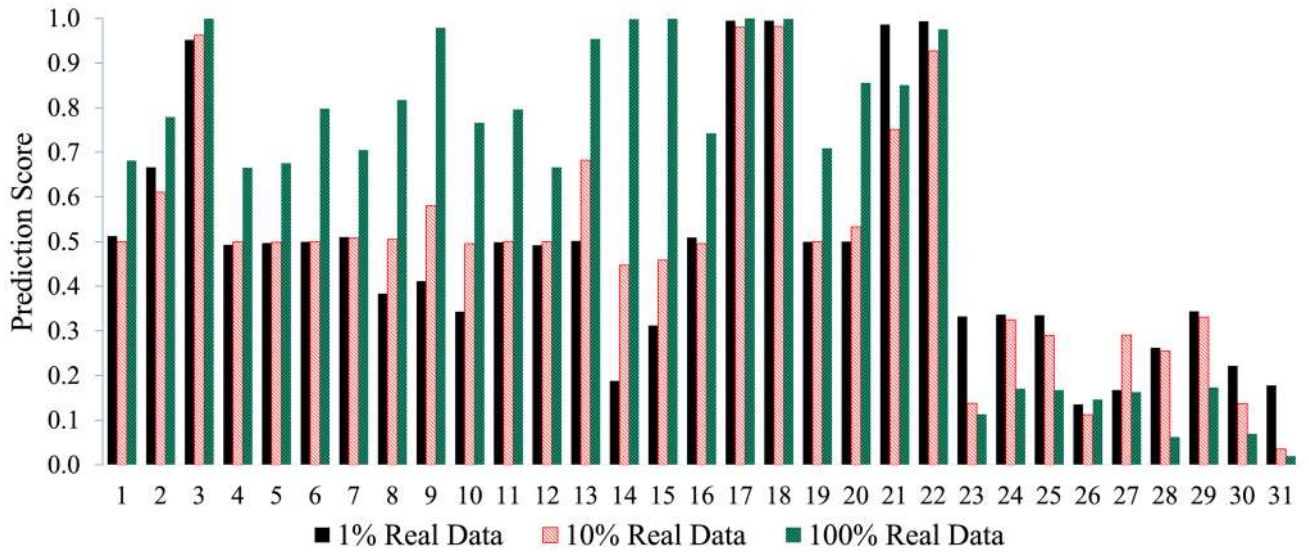

Fig. 13 Anchorage; Prediction scores for binary classification of healthy (0) and damaged (1) responses, utilizing real and HVAE-generated data at different ratios

Table 8 Comparison of binary classification performance on FOWT: results of a classifier trained on balanced data augmented using the HVAE (two imbalance ratios)

Training dataset	Biofouling				Anchorage			
	ACC	PREC	Recall	AUC	ACC	PREC	Recall	**AUC**
1% Real	0.9118	0.9405	0.9348	0.9354	0.6344	0.8556	0.5833	0.6653
10% Real	0.9322	0.9599	0.9439	0.9565	0.6720	0.8817	0.6212	0.7313
100% Real	0.9452	0.9634	0.9591	0.9624	0.8548	0.9487	0.8409	0.8973

Table 9 Component analysis of the proposed fine-tuning method for Anchorage data (1000 epochs)

Method	MSE	FID
ϵ instead of z_1 in Eq. (5)	0.4135	0.9385
w/o pre-trained model	0.0449	0.0264
HVAE	0.0054	0.0025

5 Conclusion

This paper presents the hierarchical variational autoencoder (HVAE) utilizing the pretrain-finetune training methodology to augment data in damaged scenarios for FOWTs. We have proposed an approach utilizing the HVAE concept and a diffusion process conditioned on a healthy state's latent/feature space. HVAE provides valuable data for developing predictive models that exhibit enhanced robustness to uncertainties. The proposed data generation model is evaluated comprehensively on the MNIST benchmark image dataset and compared with conventional and recently developed oversampling methods. The training and evaluation data for FOWT were obtained from simulations of the DeepCWind floater, which were performed using the OpenFast open-source platform. The following conclusions can be summarized:

- HVAE exhibits enhanced diversity and richness when generating minority MNIST samples. Generated minority samples indicate similarities in MNIST number writing style, such as tilt angle and thickness, compared to reference majority samples, despite differing semantic information. Moreover, HVAE performs better than other oversampling techniques, particularly in scenarios of significant imbalance ($\rho = 600$) within MNIST multiclass classification. The geometric mean index reflects a notable improvement of 10.2% in this scenario.
- The minority data size presents a significant challenge for augmentation and oversampling methods. An experiment is conducted to ascertain the minimum applicable size of minority data by evaluating the performance of HVAE across various minority data sets, specifically $N^- = 3$, 6, and 12 in the case of $N^+ = 625$ sea states. The findings indicate that HVAE can generate data under more sustainable conditions and with efficient fine-tuning time, utilizing a minimum of $N^- = 6$ (1%) responses for various damage statuses. In downstream experiments, it was observed that in challenging minority data scenarios, such as Anchorage, under severe randomness conditions, a minimum of 10% or more of actual damaged data is required.
- The binary classification task evaluates the downstream application for mooring systems in FOWT. Fine-tuning and data augmentation using a limited dataset of damage-associated data (1%) reduce overfitting in classifiers within substantial imbalance and randomness. However, overfitting persists in sea states closely resembling those used for fine-tuning, highlighting the need for additional sea states to enhance the HVAE's fine-tuning process. Furthermore, classification accuracy decreases in severe sea states, even with original damaged data, apparently due to pronounced nonlinear effects.
- The component analysis evaluates the effectiveness of the principal components of the HVAE technique. The findings demonstrate that training a VAE with limited or minority data, in the absence of pre-training, leads to overfitting and reduced generalization capabilities. To tackle this challenge, the latent or feature space of healthy state data serves as a robust baseline, and through innovative reparameterization, HVAE can generate and augment data more efficiently for each damaged status.

Limitation and future work Future research lines should focus on integrating physics-informed modeling and advanced regularization techniques to improve the HVAE's generative capacity, mainly when using limited minority data and in severe sea states. This approach may enable the HVAE to more accurately represent the nonlinear dynamics characteristic of these situations, thereby enhancing its efficacy for structural health monitoring in FOWTs. Future work will focus on extending the approach to further SHM-specific applications, such as bridges or buildings, confirming its generalizability and utility across diverse domains.

Acknowledgements This research is financially supported by the Spanish Ministry of Economic Affairs and Digital Transformation through

the Recovery, Transformation, and Resilience Plan, specifically in the R&D Missions within the Artificial Intelligence 2021 Programme. The funding is allocated within the framework of the IA4TES project (Artificial Intelligence for Sustainable Energy Transition) under reference number MIA.2021.M04.008. The authors also would like to acknowledge the ELKARTEK project RUL-ET funded by the Basque Government (KK-2024/00086); the "BCAM Severo Ochoa" accreditation of excellence CEX2021-001142-S / MICIN / AEI / 10.13039/501100011033; and the Basque Government through the BERC 2022–2025 program.

Author Contributions Conceptualization: H.F. and V.N.; Methodology: H.F.; Formal analysis: H.F.; Validation: H.F. and V.N. and investigation: H.F. and V.N.; Writing - original draft preparation: H.F.; Writing - review and editing: H.F. and V.N.; Funding acquisition: V.N.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This research was financially supported by the Spanish Ministry of Economic Affairs and Digital Transformation through the Recovery, Transformation, and Resilience Plan, specifically in the R&D Missions within the Artificial Intelligence 2021 Programme. Additional funding includes the IA4TES project (Artificial Intelligence for Sustainable Energy Transition) under reference number MIA.2021.M04.008, the ELKARTEK project RUL-ET funded by the Basque Government (KK-2024/00086), the "BCAM Severo Ochoa" accreditation of excellence CEX2021-001142-S / MICIN / AEI / 10.13039/501100011033, and the Basque Government through the BERC 2022–2025 program.

Data Availability All data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Williams R, Zhao F, Backwell B, Lee J, Patel A, Hutchinson M, Qiao L, Nguyen T, Lathigara A, Liang W, Fang E, Cheong J, Tan H, Francisco AM, Bui V, Nguyen T, Fiestas R, Cox R, Ruas M, Pontes De Lima J, Rabie H, Gitobu J, Ladwa R, Madan K, Shardul M, Muchiri W, Weekes N, Ogilvie C, Khinda N, Yamamura J, Stedman -MT, Wind VO (2024) GWECs global offshore wind report—global wind energy council. Technical report. Accessed January 2025. https://gwec.net/strong-2023-offshore-wind-growth-as-industry-sets-course-for-record-breaking-decade/
2. Edwards EC, Holcombe A, Brown S, Ransley E, Hann M, Greaves D (2023) Evolution of floating offshore wind platforms: a review of at-sea devices. Renew Sustain Energy Rev 183:113416. https://doi.org/10.1016/J.RSER.2023.113416
3. DNV: Floating Offshore Wind: The Next Five Years (2022). https://www.dnv.com/focus-areas/floating-offshore-wind/floating-offshore-wind-the-next-five-years/. Accessed Dec 2023
4. Chen J, Kim M-H (2021) Review of recent offshore wind turbine research and optimization methodologies in their design. J Mar Sci Eng 10(1):28. https://doi.org/10.3390/jmse10010028
5. Nava V, Ruiz-Minguela P, Perez-Moran G, Rodriguez-Arias R, Lopez-Mendia J, Villate-Martinez J-L (2019) Installation, operation and maintenance of offshore renewables. In: Renewable energy from the oceans: from wave, tidal and gradient systems to offshore wind and solar. Institution of Engineering and Technology, pp 397–424. https://doi.org/10.1049/PBPO129Ech11
6. Ford S, La Grotta A, Rasul H, Arnau R, Campo-Cossío M, Puras A, Sainz J, Rodríguez J, Fernandez J, Hunter T (2020) MooringSense-GA 851703
7. Moghadam FK, Nejad AR (2022) Online condition monitoring of floating wind turbines drivetrain by means of digital twin. Mech Syst Signal Process. https://doi.org/10.1016/j.ymssp.2021.108087
8. Longman RP, Xu Y, Sun Q, Turkan Y, Riggio M (2023) Digital twin for monitoring in-service performance of post-tensioned self-centering cross-laminated timber shear walls. J Comput Civ Eng 37(2):04022055. https://doi.org/10.1061/(ASCE)CP.1943-5487.0001050
9. Soleimani-Babakamali MH, Soleimani-Babakamali R, Sarlo R (2022) A general framework for supervised structural health monitoring and sensor output validation mitigating data imbalance with generative adversarial networks-generated high-dimensional features. Struct Health Monit 21(3):1167–1182. https://doi.org/10.1177/14759217211025488
10. Rezaniaiee Aqdam H, Ettefagh MM, Hassannejad R (2018) Health monitoring of mooring lines in floating structures using artificial neural networks. Ocean Eng 164:284–297. https://doi.org/10.1016/j.oceaneng.2018.06.056
11. Sharma S, Nava V (2024) Condition monitoring of mooring systems for floating offshore wind turbines using convolutional neural network framework coupled with Autoregressive coefficients. Ocean Eng 302:117650. https://doi.org/10.1016/J.OCEANENG.2024.117650
12. Fathnejat H, Ahmadi-Nedushan B, Hosseininejad S, Noori M, Altabey WA (2023) A data-driven structural damage identification approach using deep convolutional-attention-recurrent neural architecture under environmental effects. Eng Struct 276:1–23. https://doi.org/10.1016/j.engstruct.2022.115311
13. Altabey WA, Wu Z, Noori M, Fathnejat H (2023) Structural health monitoring of composite pipelines utilizing fiber optic sensors and an AI-based algorithm—a comprehensive numerical study. Sensors 23(8):3887. https://doi.org/10.3390/s23083887
14. Ahmadi-Nedushan B, Fathnejat H (2022) A modified teaching-learning optimization algorithm for structural damage detection using a novel damage index based on modal flexibility and strain energy under environmental variations. Eng Comput 38(1):847–874. https://doi.org/10.1007/s00366-020-01197-3
15. Fathnejat H, Torkzadeh P, Salajegheh E, Ghiasi R (2014) Structural damage detection by model updating method based on cascade feed-forward neural network as an efficient approximation mechanism. Int J Optim Civ Eng 4(4):451–472
16. Ghiasi R, Fathnejat H, Torkzadeh P (2019) A three-stage damage detection method for large-scale space structures using forward substructuring approach and enhanced bat optimization algorithm. Eng Comput 35(3):857–874. https://doi.org/10.1007/s00366-018-0636-0

17. Fathnejat H, Ahmadi-Nedushan B (2020) An efficient two-stage approach for structural damage detection using meta-heuristic algorithms and group method of data handling surrogate model. Front Struct Civ Eng 14(4):907–929. https://doi.org/10.1007/s11709-020-0628-1
18. Torkzadeh P, Fathnejat H, Ghiasi R (2016) Damage detection of plate-like structures using intelligent surrogate model. Smart Struct Syst 18(6):1233–1250. https://doi.org/10.12989/sss.2016.18.6.1233
19. Japkowicz N, Stephen S (2002) The class imbalance problem: a systematic study1. Intell Data Anal 6(5):429–449. https://doi.org/10.3233/IDA-2002-6504
20. Ai Q, Wang P, He L, Wen L, Pan L, Xu Z (2023) Generative oversampling for imbalanced data via majority-guided VAE. Proc Mach Learn Res 206:3315–3330
21. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. J Artif Intell Res 16:321–357. https://doi.org/10.1613/JAIR.953
22. Fajardo VA, Findlay D, Jaiswal C, Yin X, Houmanfar R, Xie H, Liang J, She X, Emerson DB (2021) On oversampling imbalanced data with deep conditional generative models. Expert Syst Appl 169:114463. https://doi.org/10.1016/j.eswa.2020.114463
23. Luleci F, Catbas FN, Avci O (2022) A literature review: generative adversarial networks for civil structural health monitoring. Front Built Environ. https://doi.org/10.3389/fbuil.2022.1027379
24. Catbas N (2022) A brief introduction to deep generative models for civil structural health monitoring (December)
25. Luleci F, Catbas FN, Avci O (2022) Generative adversarial networks for labeled acceleration data augmentation for structural damage detection. J Civ Struct Health Monit. https://doi.org/10.1007/s13349-022-00627-8
26. Coraça EM, Ferreira JV, Nóbrega EGO (2023) An unsupervised structural health monitoring framework based on variational autoencoders and hidden Markov models. Reliab Eng Syst Saf 231:109025. https://doi.org/10.1016/j.ress.2022.109025
27. Ma X, Lin Y, Nie Z, Ma H (2020) Structural damage identification based on unsupervised feature-extraction via variational auto-encoder. Measurement 160:107811. https://doi.org/10.1016/J.MEASUREMENT.2020.107811
28. Pollastro A, Testa G, Bilotta A, Prevete R (2022) Semi-supervised detection of structural damage using variational autoencoder and a one-class support vector machine. IEEE Access 11:67098–67112. https://doi.org/10.1109/ACCESS.2023.3291674
29. Shu X, Bao T, Li Y, Gong J, Zhang K (2021) VAE-TALSTM: a temporal attention and variational autoencoder-based long short-term memory framework for dam displacement prediction. Eng Comput. https://doi.org/10.1007/s00366-021-01362-2
30. Lee S, Jang K, Lee S, Cho H, Shin S (2023) Parametric model order reduction by machine learning for fluid–structure interaction analysis. Eng Comput 1:3. https://doi.org/10.1007/s00366-023-01782-2
31. Zhao X, Yao J, Deng W, Jia M, Liu Z (2022) Normalized conditional variational auto-encoder with adaptive focal loss for imbalanced fault diagnosis of bearing-rotor system. Mech Syst Signal Process. https://doi.org/10.1016/j.ymssp.2022.108826
32. Li L, Betti R (2023) A machine learning-based data augmentation strategy for structural damage classification in civil infrastructure system. J Civ Struct Health Monit 13(6–7):1265–1285. https://doi.org/10.1007/s13349-023-00705-5
33. Kingma DP, Welling M (2013) Auto-encoding variational Bayes. 2nd international conference on learning representations, ICLR 2014-conference track proceedings. arxiv:1312.6114
34. Havtorn JD, Frellsen J, Hauberg S, Maaløe L (2021) Hierarchical VAEs Know What They Don't Know. Technical report
35. Sohl-Dickstein J, Weiss EA, Maheswaranathan N, Ganguli S (2015) Deep unsupervised learning using nonequilibrium thermodynamics. In: Proceedings of the 32nd International Conference on Machine Learning. PMLR, pp 2256–2265. http://proceedings.mlr.press/v37/sohl-dickstein15.html
36. LeCun Y, Boser BE, Denker JS, Henderson D, Howard RE, Hubbard WE, Jackel LD (1990) Handwritten digit recognition with a back-propagation network. In: Advances in neural information processing systems (NeurIPS), Denver, CO, USA, pp 396–404.
37. OpenFAST: OpenFAST Documentation (2023). https://openfast.readthedocs.io/en/main/
38. Robertson A, Jonkman J, Masciola M, Song H, Goupee A, Coulling A, Luan C (2014) Definition of the semisubmersible floating system for phase II of OC4 (No. NREL/TP-5000-60601). National Renewable Energy Lab.(NREL), Golden, CO (United States). https://doi.org/10.2172/1155123
39. Benitz MA, Schmidt DP, Lackner MA, Stewart GM, Jonkman J, Robertson A (2014) Comparison of hydrodynamic load predictions between reduced order engineering models and computational fluid dynamics for the OC4-DeepCwind semi-submersible. In: Proceedings of the ASME 2014 33rd International Conference on Offshore Mechanics and Arctic Engineering (OMAE), San Francisco, CA, USA, Paper No. OMAE2014-23985. https://doi.org/10.1115/OMAE2014-23985
40. Gorostidi N, Nava V, Aristondo A, Pardo D (2022) Predictive maintenance of floating offshore wind turbine mooring lines using deep neural networks. J Phys Conf Ser. https://doi.org/10.1088/1742-6596/2257/1/012008
41. Ho J, Jain A, Abbeel P (2020) Denoising diffusion probabilistic models. Adv Neural Inf Process Syst 33:6840–6851
42. Heusel M, Ramsauer H, Unterthiner T, Nessler B, Hochreiter S (2017) GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In: Advances in Neural Information Processing Systems (NeurIPS), Long Beach, CA, USA, vol 30, pp 6626–6637.
43. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791
44. Huang C, Li Y, Loy CC, Tang X (2016) Learning deep representation for imbalanced classification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp 5375–5384. https://doi.org/10.1109/CVPR.2016.580
45. Kim J, Jeong J, Shin J (2020) M2m: imbalanced classification via major-to-minor translation. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE, pp 13893–13902. https://doi.org/10.1109/CVPR42600.2020.01391. https://ieeexplore.ieee.org/document/9156943/
46. Mullick SS, Datta S, Das S (2019) Generative adversarial minority oversampling. In: 2019 IEEE/CVF international conference on computer vision (ICCV), vol. 2019-October. IEEE, pp 1695–1704. https://doi.org/10.1109/ICCV.2019.00178. https://ieeexplore.ieee.org/document/9008836/
47. Branco P, Torgo L, Ribeiro RP (2017) A survey of predictive modeling on imbalanced domains. ACM Comput Surv 49(2):1–50. https://doi.org/10.1145/2907070
48. Japkowicz N (2000) The class imbalance problem: significance and strategies. In: Proceedings of the international conference on artificial intelligence (IC-AI), Las Vegas, NV, USA, vol 56. pp 111–117
49. Cui Y, Jia M, Lin T-Y, Song Y, Belongie S (2019) Class-balanced loss based on effective number of samples. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, pp 9268–9277
50. Lin YZ, Nie ZH, Ma HW (2017) Structural damage detection with automatic feature-extraction through deep learning. Comput Aided Civ Infrastruct Eng 32(12):1025–1046. https://doi.org/10.1111/mice.12313

51. Cao K, Wei C, Gaidon A, Arechiga N, Ma T (2019) Learning imbalanced datasets with label-distribution-aware margin loss. In: Advances in Neural Information Processing Systems (NeurIPS), Vancouver, BC, Canada, vol 32, pp 1567–1578.
52. Gorostidi N, Pardo D, Nava V (2023) Diagnosis of the health status of mooring systems for floating offshore wind turbines using autoencoders. Ocean Eng 287:115862. https://doi.org/10.1016/j.oceaneng.2023.115862
53. Donostia International Physics Centre: Atlas-EDR, Technical Documentation (2024). https://dipc.ehu.eus/en
54. Dadras Eslamlou A, Huang S (2022) Artificial-neural-network-based surrogate models for structural health monitoring of civil structures: a literature review. Buildings 12(12):2067. https://doi.org/10.3390/BUILDINGS12122067

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.