Unsupervised Feature Engineering in Imbalanced and Unlabeled Mooring Monitoring Data for Augmentation and Translation by Hierarchical Variational Approach

HAMED FATHNEJAT and VINCENZO NAVA

ABSTRACT

The stability of mooring systems in floating offshore wind turbines (FOWTs) is essential, as any deterioration affects platform performance and overall efficiency. Machine-learning monitoring of FOWTs requires comprehensive dynamic response data derived from various mooring systems. Data collection should encompass a variety of health states, operational scenarios, and metocean conditions. Data is scarce in practice, especially concerning damage-associated information. Additionally, the existing imbalanced dataset lacks clear labels, making it difficult to differentiate between healthy (majority) and damage-associated (minority) instances. We propose a novel deep generative model (DGM) to efficiently deal with the challenge of unsupervised labeling of imbalanced data in the initial stage. It will subsequently be used to augment the labeled data from the previous stage and generate damage-associated data for a new (target) mooring system based solely on the healthy records of that system. The conditional hierarchical variational autoencoder (CHVAE) is developed using a diffusion probabilistic architecture during a pretrain-finetune training procedure. During the pre-training stage, trained on imbalanced, unlabeled data, we employed an unsupervised feature engineering approach in the latent space created to identify the healthy (majority) and damaged (minority) data. In the fine-tuning stage, the nonlinear relationship between healthy and minority-damaged responses of a mooring system is obtained, considered as the source domain across diverse sea states. Subsequently, by employing the healthy data from the target mooring system, the CHVAE can generate real-scale damaged responses of the system across diverse operational and environmental conditions. An analysis is conducted to evaluate the similarity between the simulated records produced by OpenFast on the OC4-DeepCWind FOWT benchmark and those generated by the CHVAE framework, utilizing visual, statistical, and behavioral methods. Furthermore, the proposed DGM's performance is first assessed using MNIST benchmark image dataset to evaluate its effectiveness for unsupervised labeling and data augmentation. The generated records for unobserved random sea states closely resemble real-world dynamic behaviors during downstream binary classification, illustrating the effectiveness and versatility of the proposed DGM, CHVAE.

*

INTRODUCTION

Floating offshore wind turbines (FOWTs) are increasingly used to capture wind energy in deep-water fields. The stability and integrity of mooring systems are essential for safe and efficient operation, as any deterioration may compromise structural reliability and energy production efficiency [1]. Key indicators of mooring health include fairlead tension responses, which demonstrate the dynamic behavior of mooring lines under environmental and operational loading [2].

Recent advancements in structural health monitoring (SHM) have increasingly utilized data-driven approaches, employing machine learning techniques to analyze intricate dynamic responses. The lack of labeled damage data frequently constrains the effectiveness of these approaches. In practice, damage events are infrequent, and the data are often unlabeled or noisy, leading to highly imbalanced datasets primarily characterized by healthy conditions [3, 4]. This issue is particularly pronounced in floating systems, owing to the intricate interactions between environmental forces and structural dynamics. Recent approaches, including oversampling and cost-sensitive learning, have enhanced data availability [5].

Recent research has utilized DGMs, such as generative adversarial networks (GANs) and VAEs, to generate damage data and equilibrate SHM datasets [4]. Advanced methods focus on domain adaptation, facilitating the generation of damaged data for target systems using only healthy data. This developing field facilitates zero-shot transfer learning for SHM [6].Furthermore, VAEs have been extensively utilized for unsupervised feature learning and anomaly detection. In SHM applications, these methods facilitate the extraction of latent representations of normal behavior, which can be utilized to identify anomalies or pseudo-label infrequent events [7, 8].

We propose a two-stage framework integrating unsupervised learning and generative modeling for damage identification and augmentation. This study examines fairlead tension time series data from the OC4-DeepCWind semi-submersible platform, a recognized reference system for analyzing floating wind turbines.

Initially, a Variational Autoencoder (VAE) [9] is trained to acquire latent representations of the unlabeled tension data, encompassing both the majority (healthy) and minority (damaged) class data. Unsupervised feature engineering in this latent space allows for the identification of outliers linked to the damaged state (minority data) in comparison to the majority (healthy), facilitating the pseudo-labeling of the minority damaged state data.

In the second stage, we present a deep generative model (DGM), a conditional hierarchical VAE (CHVAE) that enhances the pseudo-labeled damage data and enables domain translation. Our DGM can synthesize realistic damage responses for a new mooring system using only its healthy recordings. The zero-shot augmentation capability is crucial

Hamed Fathnejat, Post-doc Fellow, Email: hfathnejat@bcamath.org Basque Center for Applied Mathematics, 14 Mazarredo Ave, Bilbao, 48009, Biscay, Spain, ORCID: 0000-0002-1146-8384

Vincenzo Nava, Associate Professor, Email: vincenzo.nava@polito.it Department of Environment, Land and Infrastructure Engineering (DIATI), Politecnico di Torino, Corso Duca degli Abruzzi, 24, Turin, 10129, Italy, ORCID: 0000-0002-3207-3517

for generalizing SHM models to novel deployments.

We initially evaluate the CHVAE on the MNIST image benchmark [10] to validate its effectiveness and versatility in unsupervised labeling and augmentation. Subsequently, as the source system, we utilize datasets created by OpenFAST ([11]) utilizing the OC4 platform's numerical model from the DeepCwind project ([12]). We used an alternate mooring configuration for the target system, which had equal pretension but varied mass per unit length and axial stiffness (Orcaflex handbook for studless chains, [13]).

METHODOLOGY AND CASE STUDIES

PRE-TRAINED VAE

The VAE requires denormalization settings for practical training to rebuild real-scale data, especially time-series data. Damage-associated target mooring data and denormalization parameters are unknown, making domain translation challenging—a second decoder, conditioned on denormalization parameters, sampled latent space, and estimated parameters. The encoder latent space depends on denormalization parameters. The initial decoder reconstructs normalized input data during training, while the second conditional decoder calculates denormalization parameters from the conditional latent space.

Pre-training is performed on unlabeled data, including a majority class (Healthy) and a minority class (Damaged-state). The trainable weights of the proposed VAE are optimized based on Eq. (1):

$$\begin{aligned} \mathrm{ELBO}_{\theta,\theta',\phi}(X^+, d_1) = \mathbb{E}_{q_\phi}\left[\log p_\theta(X^+ \mid z_1)\right] + \mathbb{E}_{q_\phi}\left[\log p_{\theta'}(d_1 \mid z_1, d_1)\right] \\ - \mathrm{KL}\left[q_\phi(z_1 \mid X^+, d_1) \parallel p_\theta(z_1 \mid d_1)\right] \end{aligned} \tag{1}$$

Eq. (1) uses X^+ to represent normalized healthy data from the source mooring and d_1 to represent its denormalization matrix. After training, the latent Matrix was retrieved to perform unsupervised majority and minority class characterization. A maximum window of size Numframes in each record or digit was used with a stride of one to calculate the maximum for each row. A matrix of NumRecords and latent dimension dimensions was created. A two-dimensional t-distributed Stochastic Neighbor Embedding (t-SNE) transformation was done to this Matrix to reveal the data's low-dimensional structure, which is label-free. The main class-discriminative variance appears in the initial t-SNE component. This dimension simplifies representation while preserving discriminating information. Data clustering patterns were analyzed using unsupervised feature engineering. The latent space structure grew organically and directed pseudo-labeling of dominant and aberrant patterns, preparing for generative modeling.

FINE-TUNED VAE

Data generation for target mooring damage severities is controlled by a conditionally fine-tuned VAE trained on source mooring data. The damage severity/state, S, influences the encoder and decoder to create a minority latent space and produce minority damage-associated data. Inspired by diffusion, the conditional hierarchical VAE (CHVAE) creates two latent variables, z_1 and z_2, with two stochastic layers. We use a variational approximation to the true posterior and a bottom-up formula for the inference model to expand the loss function of the standard VAE, resulting in Eq. (2):

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

According to Eq. (2), x^- represents minority-damaged state data, z_1 represents the healthy or baseline condition, and z_2 represents the damaged state.

Based on the assumptions used in the reparameterization of the pre-trained VAE (Eq. (1)), $q_\phi(z_1 \mid z_2)$ follows a conventional normal distribution since z_1 and z_2 have the same dimension (D). Consequently, the second KL divergence factor in Eq. (2) remains constant during fine-tuning. Thus, analyzing the conditional fine-tuned VAE yields Eq. (3):

$$\begin{aligned}
L_{\text{CHVAE}}(x^-, S; \phi, \theta) = \mathcal{L}_{\text{recon},\theta}(x^-, \hat{x}^-) \\
+ \beta \times \left(-\text{KL}\left[q_\phi(z_2 \mid x^-, S) \parallel p_\theta(z_2 \mid z_1, S)\right]\right) \\
+ \text{FID}_{\text{loss}}
\end{aligned} \tag{3}$$

The term $\mathcal{L}_{\text{recon}}$ in Eq. (3) quantifies the discrepancies between the input x^- and its reconstruction $\hat{x}^-$. The reconstruction error function in a Variational Autoencoder (VAE) is influenced by the data characteristics and the activation function used in the decoder's final layer. Binary Cross-Entropy (BCE) is frequently employed for binary or binarized data, especially in conjunction with a sigmoid activation function. Conversely, mean squared error (MSE) is more appropriate for continuous data, such as time-series data, often utilizing tanh or linear activation functions. This method improves model performance by aligning reconstruction loss with the data's characteristics and the task's requirements.

The Fréchet Inception Distance (FID) loss is calculated using Equation (4).

$$\text{FID}_{\text{loss}} = \|\mu_{x^-} - \mu_{\hat{x}^-}\|_2^2 + \text{Tr}\left(C_{x^-} + C_{\hat{x}^-} - 2\left(C_{x^-} C_{\hat{x}^-}\right)^{1/2}\right) \tag{4}$$

In Eq. (4), 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 FID serves as an index for monitoring the fine-tuning process, where lower FID values signify greater similarity between the compared datasets. The training framework of the proposed DGM, CHVAE, is illustrated in Fig. 1.

MNIST DATASET

MNIST benchmark image dataset ([10]) classifies digits 0 to 4 as the majority classes, with 6,000 samples in each class, totaling 30,000 samples. The minority classes

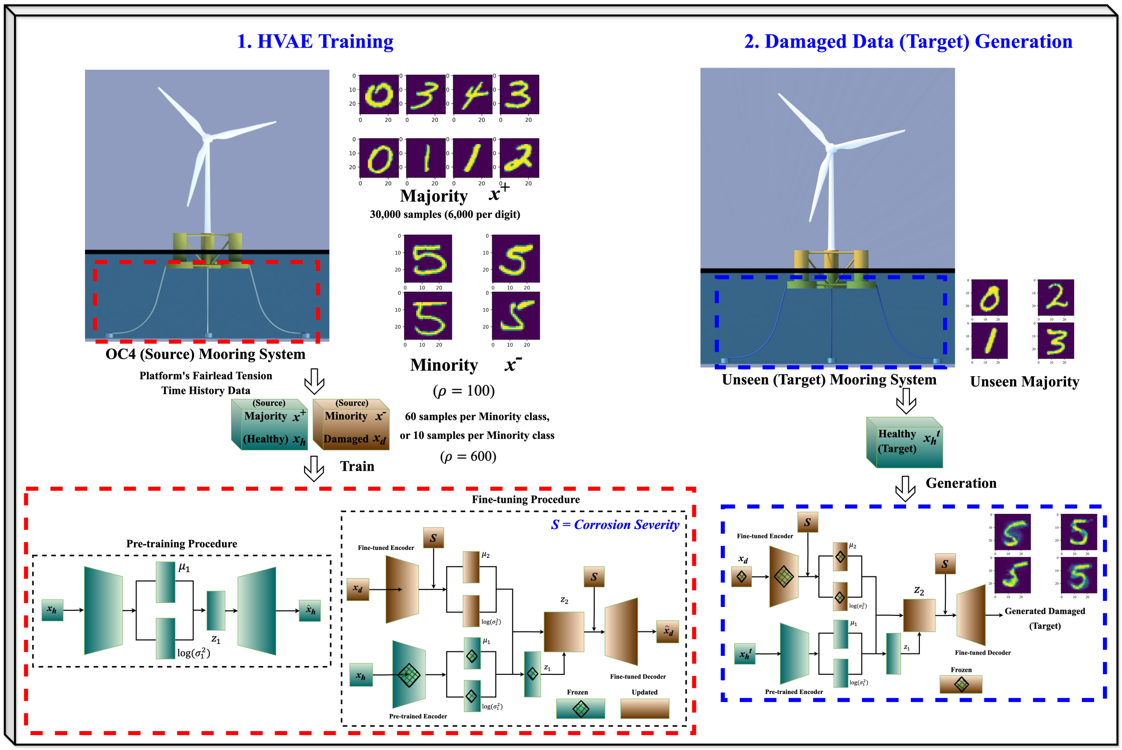


Figure 1. The illustration of the proposed minority (damaged) data generation procedure for an unseen minority or damaged records of the target FOWT mooring system.

(digits 5-9), with 300 samples (60 per class) and 50 samples (10 per class), had imbalance ratios of $\rho = 100$ and $\rho = 600$, respectively. The downstream classifier uses a two-layer fully connected neural network, similar to the architecture used in [14]. The CNN-based VAE architecture, trained on the majority (x^+), uses five 2D CNN layers with a kernel size of (3×3) for the encoder. The layers include 32, 64, 128, 256, and 512 filters and an input shape of (Batch size, 28, 28, 1). The stride is 1, except for the final CNN layer, which uses 2. All levels except the output layer use ReLU activation. MLP layers create latent. The decoder uses five Conv2DTranspose layers to match encoder features. Fine-tuned convolutional neural network-based variational autoencoder architecture: Encoder: Input01(Batch size, 28, 28, 1), Conv2D(32, (3, 3), stride=(1, 1)), 64, (3, 3), stride=(2, 2), 64, (3, 3), stride=(1, 1), 128, (3, 3), stride=(1, 1)). Decoder: Input01(z_2), Input02(S), MLP($14 \times 14 \times 128$), Reshape(14, 14, 128). Conv2DTranspose(128 (3, 3), strides=(1, 1)), 64 (3, 3), strides=(1, 1)), CONV2DTranspose with 64 filters, kernel size (3, 3), and stride (1, 1); with 32 filters, kernel size (3, 3), and stride (2, 2). All levels except the output layer use ReLU activation. The final decoder layer uses sigmoid activation. The reconstruction loss function is Binary Cross-Entropy (BCE), and the optimizer is Adam, with a 0.001 learning rate. The pre-training batch size is 32. Batch sizes of 300 and 50 were optimized for $\rho = 100$ and $\rho = 600$. For both the pre-training VAE and fine-tuning CHVAE, Equations (1) and (3) use a β value of 5×10^{-4}.

FOWT DATASET

The tension records in the semisubmersible floating system fairleads, shown in Fig. 2, are crucial for assessing mooring system integrity. Tension time-series training, validation, and testing responses are generated under various randomly sampled sea state conditions. Excitation variability is shown by wave height (H_s) between 1 and 7 meters, peak period (T_p) between 8 and 15 seconds, wind speed (V) between 1 and 10 m/s, and current speed (C) between 0.5 and 1.5 m/s. The source and target systems' mooring line characteristics are shown in Table I.

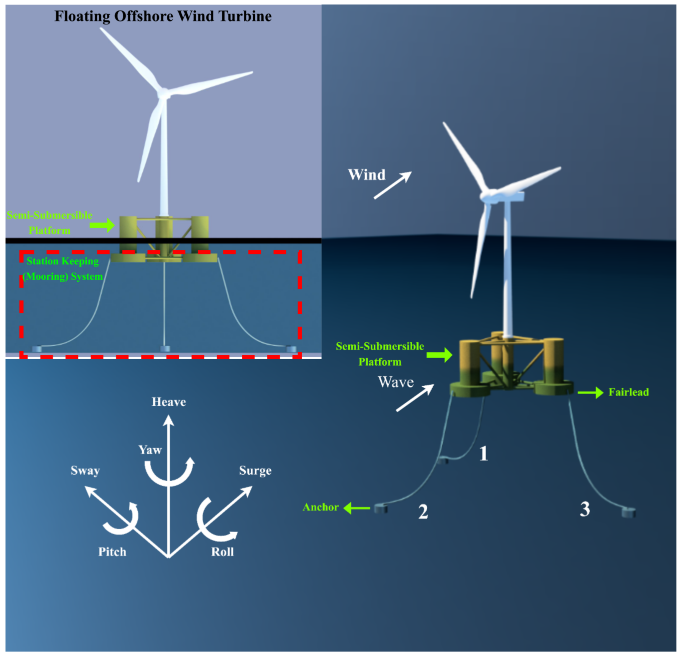


Figure 2. Schematic sketch of the OC4-DeepCwind semisubmersible FOWT.

TABLE I. The mooring systems' properties; a three catenary line system.

Property	Source (OC4)	Target (Orcaflex Studless Chain)	Unit
Number of lines	3	3	–
Segments per line	20	20	–
Line diameter	7.66×10^{-2}	8.50×10^{-2}	m
Mass/length	1.13×10^{2}	1.44×10^{2}	kg/m
Axial stiffness (EA)	7.54×10^{8}	6.17×10^{8}	N
Unstretched length	8.35×10^{2}	8.25×10^{2}	m

We examined the corrosion impact by decreasing the axial stiffness (EA (N)) and mass per unit length ($Mass\ density$ (kg/m)) of all three mooring lines, which degraded differently in the MoorDyn module. In this experiment, mooring lines 1 and 2 show degradation of 3%, 5%, and 7%, whereas line 3 shows 5%, 7%, and 10% reductions in EA and $Mass$ density. This corrosion type was chosen to represent mooring line degradation, as the Orcaflex manual suggests studless chains have a lower safety factor [13]. The dataset includes a 1500-second fairlead tension time series (7500 steps) recorded at 5 Hz, three features (FairTen1, FairTen2, and FairTen3), and three mooring lines (Fig. 2). Trial and error determined the VAE layer input frame size of 50. Thus, the input data structure is $(batchsize, 50, 3, 1)$. The source mooring system (OC4) is fine-tuned using healthy and 40% damaged data to improve the CHVAE model. After fine-tuning, the model creates damaged target mooring system responses using healthy data from the same system. Figure 1 visually represents the proposed domain translation framework. The VAE pre-training encoder and decoder architecture is: Encoder: Input01 (Batch size: 50, 3, 1), Conv2D (32), 3×Conv2D (64), MLP (32); Decoder: Input(z_1), 2×Conv2DTranspose(64), 32). The encoder and decoder architecture for CHVAE fine-tuning is: The encoder uses Input01 (Batch size: 50, 3, 1), Conv2D (32), 3×Conv2D (64), MLP (32), and Input02 (S), whereas the decoder uses Input01 (z_2), Input02 (S), Conv2DTranspose(128), and 3×Conv2DTranspose(64). The convolution window is 2x2, with a 1x1 stride and identical padding. All levels except the output layer use ReLU activation. The last decoder layer uses tanh for activation. MSE is the reconstruction loss function, while Adam, with a learning rate of 0.001, is the optimizer. Pre-training and fine-tuning batches are 32 and 4650. A β value of 5×10^{-4} is used for VAE pre-training and CHVAE fine-tuning.

RESULTS AND DISCUSSION

UNSUPERVISED LATENT SPACE SEPARATION

The first dimension of the t-SNE projection of two-dimensional Max-Window latent features from 875 sea states is shown in Fig. 3. Each point represents a sea state. The first 625 sea states represent healthy circumstances, whereas the next 250 suggest damage. Healthy samples dominate the t-SNE projection, while damage-associated samples are segregated in the 1D embedding space. The emerging structure, found without labels, confirms that the VAE latent space represents health-relevant dynamics and that max-pooled latent features preserve discriminatory information. We use unsupervised separation to support our pseudo-labeling technique and the proposed CHVAE's augmentation and translation tasks.

QUALITATIVE ASSESSMENT

Fig. 4 displays qualitative outcomes of MNIST data augmentation utilizing CHVAE. Figure 4 shows that for each digit in the majority class (0, 1, 2, 3, and 4), all digits in the minority category (5, 6, 7, 8, and 9) may be generated. The definition of S generates minority digits 5–9 by changing the input digit. Additionally, Fig. 4 shows stylistic similarities between minority data and majority genuine data in aspects like tilt angle, thickness, and structure, despite semantic disparities. Keeping tilt and thickness consistent (handwritten style), minority digits seem like majority values. This consistency shows that CHVAE can capture basic data features for realistic augmentation.

See Fig. 5 for the target mooring system's generated and original damaged tension data in a random sea state. Two levels of corrosion severity are examined in the target mooring system: mild (3-5%) and high (7-10%). Fig. 5 (a) shows tension time se-

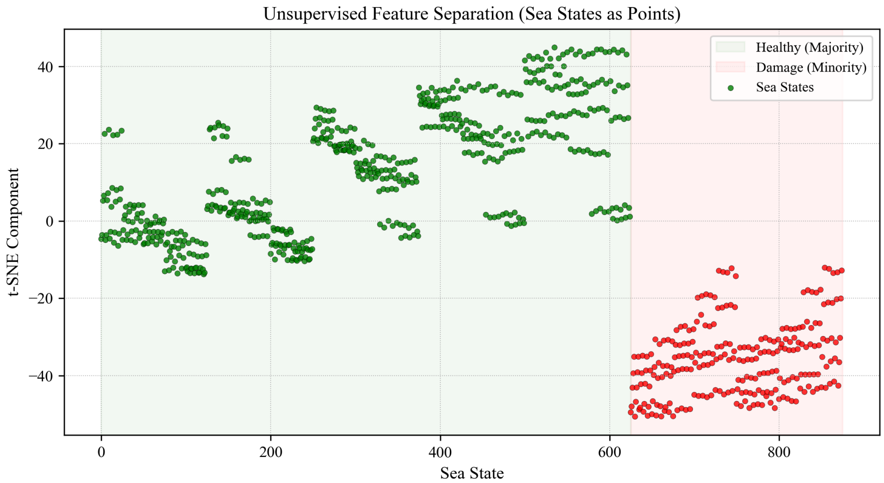
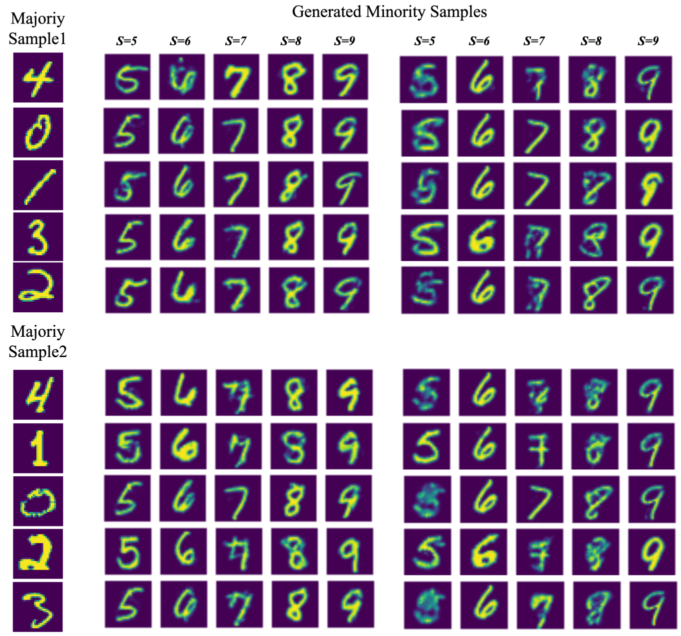


Figure 3. Scatter plot of the first dimension of the 2D t-SNE projection of latent features from fairlead tension responses across 875 sea states. Each dot represents one sea state. The green region (samples 1–625) corresponds to healthy states, and the red region (samples 626–875) corresponds to damage-associated states. The clear separation highlights the capability of unsupervised feature engineering to distinguish health conditions without supervision.

ries for different corrosion severities and healthy (non-damaged) and damaged records. Fig. 5 (b) contains a zoomed-in smaller dataset time series. The created damaged records closely resemble the actual ones at the stated severities after zooming in, indicating that the CHVAE can generate real-scale stress patterns for different damage levels. Fig. 5 Lower frequency power spectral densities (PSDs) when tension dynamics dominate. The PSDs align the original data over several frequency domain severities, suggesting that the CHVAE preserves spectral properties.

BEHAVIOURAL ASSESSMENT

The CHVAE generates significant results for MNIST at $\rho = 100$, as seen in Table II. In situation of considerable imbalance ($\rho = 600$), the GM term outperforms MGVAE by 3.7. CHVAE works under severe imbalance, according to the findings.



(a) (ρ = 100) (b) (ρ = 600)

Figure 4. Visualization of MNIST data generation by CHVAE.

TABLE II. Comparison of CHVAE classification performance on MNIST under different imbalance ratios (IR).

IR	$\rho = 100$			$\rho = 600$		
Methods	**B-ACC**	**ASCA**	**GM**	**B-ACC**	**ASCA**	**GM**
MGVAE [14]	85.0 ± 0.2	84.6 ± 0.2	83.2 ± 0.2	65.4 ± 1.0	64.4 ± 1.1	53.4 ± 1.1
CHVAE	**90.5 ± 0.2**	**90.3 ± 0.2**	**90.1 ± 0.2**	**67.3 ± 0.4**	**67.2 ± 0.4**	**57.1 ± 1.1**
Fully-Balanced MNIST ($\rho = 1$)						
	B-ACC	**ASCA**	**GM**			
	98.5 ± 0.0	98.5 ± 0.0	98.5 ± 0.0			

The classifier design for real-time target mooring monitoring uses a fully connected neural network with three components: C_1 = MLP [Input dim – 256], C_2 = MLP [256 – 128], and C_3 = MLP [128 – 1]. For the last dense layer, the Adam optimizer optimizes a sigmoid activation function with binary cross-entropy loss function at a 1×10^{-3} learning rate. Sigmoid output predicts 0 (healthy) and 1 (damaged). For optimal performance, use an input shape of $(6, 1250 \times 3)$ for each record. We train the classifier for $25epochsin200$ batches. We normalize all three features using Robust Scaling because it handles outliers best and improves categorization. The classifier was evaluated with 31 unseen stress recordings from different damage severities, operating sea conditions, and wave seeds. 22 records were damaged, 9 intact. See Table III for classification findings.

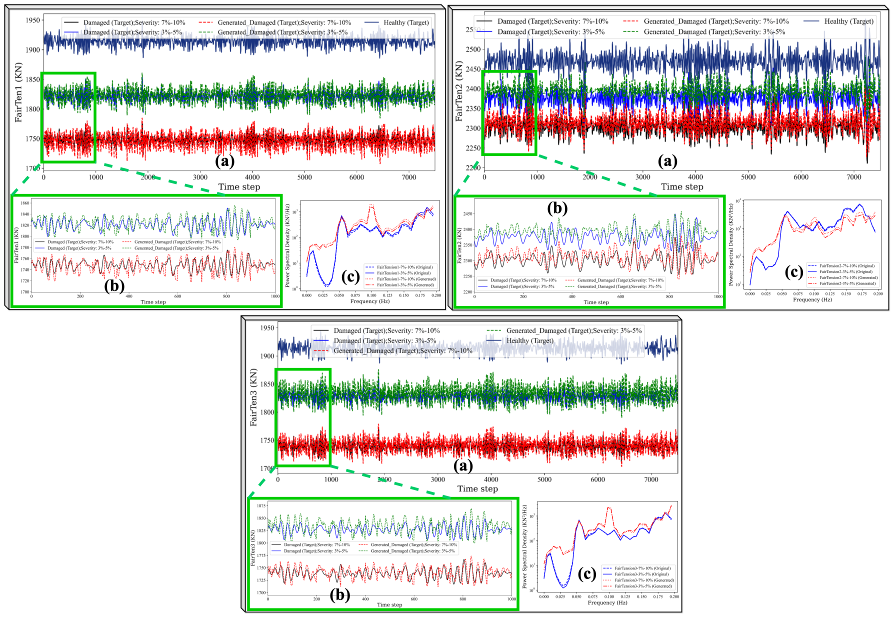


Figure 5. Typical sample; (a) the unseen fairlead tensions (Mooring lines 1, 2, and 3) of platform coupled with target mooring system under healthy and damaged operational state when $V = 7.75$,$H_s = 1.0$,$T_p = 15.0$,$C = 1.25$. (b) The first 1000 time steps of real and approximated damaged record. (c) Power spectral density of tension record.

TABLE III. The target mooring system classification metrics.

Classification Type	**Accuracy**	**Precision**	**Recall**	**AUC (PR)**
With Blind Domain Translation	0.9408	1.00	0.9167	0.9796
Without Blind Domain Translation	0.7096	1.00	0.5909	0.8446

Table III shows that the classifier trained on generated damage-associated records accurately identifies the most damaged data frames, significantly outperforming the target mooring system's lack of information.

CONCLUDING REMARKS

This study confirms that latent dynamics summarized by a max moving window and embedded by 2D t-SNE maintain enough structure to distinguish healthy and damage-associated sea states unsupervised. The first dimension of the t-SNE projection alone captures the majority of the variance, allowing for simplified pseudo-labeling in imbalanced and unlabeled mooring monitoring data. This paper uses The pretrain-finetune training framework to introduce the conditional hierarchical variational autoencoder (CHVAE) data-generating model. CHVAE uses domain translation to supplement source mooring system damage data across operational sea states and generate labeled target mooring system damage records. This framework strengthens damage detection systems in varied operational settings and promotes mooring system knowledge transfer. According to the research, a shallow architecture classifier can label unseen, real-time data from the target mooring system due to downstream record performance.

ACKNOWLEDGMENT

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.

REFERENCES

1. Lauria, A., P. Loprieno, F. Rizzo, A. Severini, D. Foti, E. Leone, A. Francone, and G. Tomasicchio. 2024. "On the effects of wind and operating conditions on mooring line tensions for floating offshore wind turbine," *Applied Ocean Research*, 152:104197.
2. Gräfe, M. J., V. Pettas, N. Dimitrov, and P. W. Cheng. 2024. "Machine learning based virtual load sensors for mooring lines using motion and lidar measurements," *Wind Energy Science Discussions*, 2024:1–29.
3. Prawin, J. 2025. "Deep learning neural networks with input processing for vibration-based bearing fault diagnosis under imbalanced data conditions," *Structural Health Monitoring*, 24(2):883–908.
4. Luleci, F. and F. N. Catbas. 2023. "A brief introductory review to deep generative models for civil structural health monitoring," *AI in Civil Engineering*, 2(1):9.
5. Altalhan, M., A. Algarni, and M. T.-H. Alouane. 2025. "Imbalanced Data problem in Machine Learning: A review," *IEEE Access*.
6. Kiranyaz, S., O. C. Devecioglu, A. Alhams, S. Sassi, T. Ince, O. Abdeljaber, O. Avci, and M. Gabbouj. 2024. "Zero-shot motor health monitoring by blind domain transition," *Mechanical Systems and Signal Processing*, 210:111147.
7. Wan, H.-P., Y.-K. Zhu, Y. Luo, and M. D. Todd. 2025. "Unsupervised deep learning approach for structural anomaly detection using probabilistic features," *Structural Health Monitoring*, 24(1):3–33.
8. Okur, F. Y., A. C. Altunişik, and E. Kalkan Okur. 2025. "A Novel Approach for Anomaly Detection in Vibration-Based Structural Health Monitoring Using Autoencoders in Deep Learning," *Structural Control and Health Monitoring*, 2025(1):5602604.
9. Kingma, D. P., M. Welling, et al. 2013, "Auto-encoding variational bayes," .
10. LeCun, Y., B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel. 1989. "Handwritten digit recognition with a back-propagation network," *Advances in neural information processing systems*, 2.
11. OpenFAST. 2023, "OpenFAST Documentation," .
12. Robertson, A., J. Jonkman, M. Masciola, H. Song, A. Goupee, A. Coulling, and C. Luan. 2014. "Definition of the semisubmersible floating system for phase II of OC4," Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States).
13. Orcina. 2022, "Chain: OrcaFlex User Manual. Version 12.1," .

14. Ai, Q., P. Wang, L. He, L. Wen, L. Pan, and Z. Xu. 2023. “Generative oversampling for imbalanced data via majority-guided VAE,” in *International Conference on Artificial Intelligence and Statistics*, PMLR, pp. 3315–3330.