Future Cities and Environment
ISSN 2363-9075 Volume 12 (2026)
© Cerebration Science Publishing
https://doi.org/10.70917/fce-2026-010

Article

Field-Validated Digital Twin–Enabled Structural Health Monitoring for Offshore High-Pile Wharves

Seyed Reza Samaei [1,*] and James Riffat [2]
[1] Department of Marine Structures, Faculty of Engineering, Science and Research Branch, Islamic Azad University, Tehran 1477893855, Iran. P.O. Box: 14515-775.
[2] World Society of Sustainable Energy Technologies, Nottingham, NG11 6AA, United Kingdom
* Correspondence author: samaei@srbiau.ac.ir

Abstract: Offshore high-pile wharves operate under harsh, variable loads that can mask the earliest signs of deterioration. We present a closed-loop structural health monitoring (SHM) framework that couples multimodal machine/deep learning with a physics-based digital twin (DT). A year-long field deployment at an operating high-pile wharf in Western Asia instrumented six piles with triaxial accelerometers and strain gauges (GPS-synchronized, 100 Hz). After quality control and environmental/operational variability normalization, an unsupervised VAE produces an anomaly score that is fused with a CNN–GNN classifier to yield window-level damage probabilities and pile-by-elevation localization. The DT assimilates these diagnostics through weighted least squares and an EKF, with measurement covariance derived from model uncertainty; "what-if" simulations quantify the structural consequence of candidate repairs. Under chronological splits with leave-one-pile-out, the framework achieves F1 = 0.95 and AUC-ROC = 0.98, with false-alarm rate ≈ 2–5 % and median detection latency ≈ 45 s. Spatial heatmaps consistently pinpoint the affected elevation, while online updating reduces mean frequency bias from ≈ 4.8 % to ≈ 0.9 % and increases mode-shape MAC from ≈ 0.81 to ≥ 0.92. First-mode shifts of ≈ 3–4 % observed in the field are reproduced by DT "stiffness-restoration" scenarios, enabling risk-informed maintenance planning. Compared on identical normalized windows, the proposed method outperforms classical vibration-based indices in both discrimination and false-alarm control. The results demonstrate a scalable, explainable pathway to predictive asset management for critical maritime infrastructure. Beyond structural safety, the proposed framework contributes to sustainable port operation by enabling condition-based maintenance, reducing unnecessary inspections, extending asset service life, and lowering the embodied carbon associated with premature repair or replacement.

Keywords: Structural health monitoring (SHM); high-pile wharf; digital twin; variational autoencoder (VAE); CNN–GNN classifier; damage localization; environmental/operational variability (EOV); extended Kalman filter (EKF); modal updating; uncertainty quantification.

1. Introduction

High-pile wharves are central to the operation of many ports. They carry the weight of petrochemical exports, international trade, and routine port activities, yet they must do so while facing harsh marine conditions day after day. Constant loading from waves, currents, berthing forces, and heavy equipment gradually reduces their strength. With time, this slow degradation can turn into a serious safety risk if it is not carefully monitored and managed. Inspections remain the conventional way of safeguarding these structures, but such checks are costly, carried out only at intervals, and often miss the earliest signs of damage. They offer engineers a snapshot rather than a continuous record, which means problems may only be recognized after they have already progressed. This reality has

Copyright: © 2026 by the authors

driven the search for monitoring methods that can follow the behavior of the structure in real time and give early warnings when changes begin to appear.

Structural health monitoring (SHM) provides a framework for this type of continuous observation. The goal is straightforward: to detect and interpret signs of damage as they happen, so that maintenance can be proactive rather than reactive. In recent years, advances in data-driven technology have reshaped what SHM can achieve. Machine learning (ML) and deep learning (DL) models are able to work through large volumes of sensor data and reveal patterns that might otherwise be invisible, while digital twins (DTs) create virtual counterparts of physical assets that are updated with live information. The combination of these tools makes it possible not only to monitor but also to simulate and even anticipate how a structure will behave under changing conditions. Although SHM has been studied extensively for bridges, offshore platforms, and wind turbines, real-world field studies on offshore high-pile wharves are still very limited. Much of the existing knowledge comes from laboratory experiments or numerical models, which cannot fully reproduce the complexity of an operating marine environment. This gap makes it difficult for engineers and port authorities to rely on data-driven systems when planning inspections or scheduling maintenance.

In contrast, the present study advances the state of the art by delivering a fully closed-loop SHM–digital twin system validated through a 12-month field deployment on an operating offshore high-pile wharf. Unlike recent DT–SHM studies that focus on either data-driven inference or model updating in isolation, this work integrates multimodal deep learning, uncertainty-aware data assimilation, and decision-oriented what-if simulations within a single operational framework. This combination enables not only detection and localization of damage, but also physically interpretable prediction of structural consequences under real operating conditions.

Recent digital twin–enabled SHM frameworks have demonstrated promising results for bridges and large-span civil structures, primarily at numerical or laboratory scales. For instance, (Wang et al.2023) proposed a hierarchical deep-learning-driven digital twin for cable-dome structures, while (Li et al. 2022) and related studies in Structural Health Monitoring emphasized response estimation and damage classification under controlled conditions. However, these frameworks typically rely on short-term datasets, lack sustained environmental and operational variability, and do not close the loop between machine learning diagnostics and continuous physics-based model updating. Compared with the authors' earlier work on jack-up platform legs (Samaei & Riffat, 2025), the present study extends the methodology in three fundamental ways. First, it addresses a structurally and operationally distinct asset class—offshore high-pile wharves—characterized by distributed pile groups, complex soil–structure interaction, and continuous port operations. Second, the digital twin in this study is updated online using field-derived uncertainty from multimodal learning outputs, rather than periodic or batch-based calibration. Third, the validation relies on long-term in-situ monitoring under real environmental and operational variability, moving beyond the episodic or campaign-based data used in prior studies. These differences establish the present work as a substantive methodological and application-level advance rather than an incremental extension.

The present work responds to this need by proposing an integrated SHM framework tailored to offshore wharf piles. The system links multimodal ML/DL models with a continuously updated digital twin and is supported by a dense network of sensors deployed in the field. Long-term measurements are used to test its capacity to identify, localize, and track structural deterioration under real operating conditions. The findings show that this approach offers clear improvements over classical vibration-based indices and points toward more reliable and cost-effective management of offshore infrastructure. From a future cities and sustainability perspective, port infrastructures are critical interfaces between urban economies, global supply chains, and coastal environments. Improving their reliability and extending their service life directly reduces material consumption, embodied carbon, and disruption-related emissions. Digital twin–enabled SHM frameworks, such as the one proposed in this study, support these goals by enabling predictive maintenance and resource-efficient asset management, aligning offshore infrastructure monitoring with broader sustainability objectives.

2. Literature Review

2.1. Degradation mechanisms and monitoring needs in high-pile wharves

High-pile wharves endure persistent wave-induced loads, ship berthing impacts, corrosion, scour around piles, and fatigue at welded connections—conditions that differ markedly from land-based structures. Recent marine studies demonstrate that pile damage can be subtle and distributed, often initiating below waterline where direct access is limited; reliable monitoring therefore hinges on vibration/strain features that remain sensitive under wave excitation and operational variability (Lecieux et al., 2019; Samaei et al., 2021; Wang et al., 2022; Zhu et al., 2023; Zheng et al., 2024; Xu et al., 2024). Experiments on framed bents and wharf pile groups show that stiffness loss, period elongation, and high-order response statistics under waves provide early indicators for localization and severity assessment (Zheng et al., 2024; Zhu et al., 2023; Wang et al., 2022; Xu et al., 2024).

2.2. SHM foundations: from vibration-based to data-driven paradigms

Classical SHM organizes the diagnostic task from detection through localization, quantification, and prognosis (Riffat et al., 2025; Farrar & Worden, 2007; Worden & Manson, 2007; Samaei et al., 2021). Foundational surveys and monographs established the use of modal features (frequencies, mode shapes/curvatures, flexibility) and model-updating for damage inference (Doebling et al., 1996; Brownjohn, 2007; Figueiredo et al., 2011; Beck & Katafygiotis, 1998). A persistent challenge is environmental/operational variability (EOV)—temperature, sea state, berthing loads—whose confounding effects can mask small changes due to damage. Cointegration-based normalization and related time-series tools are effective at removing nonstationary trends and restoring feature sensitivity (Cross et al., 2011; Shi et al., 2018). For marine assets exposed to waves and tides, robust EOV handling is a prerequisite for reliable long-term alerts.

2.3. Machine and deep learning for SHM: vibration and vision

The adoption of machine learning (ML) and deep learning (DL) has shifted SHM from hand-crafted indicators to learned representations from raw signals or images (Riffat et al., 2025; Worden & Manson, 2007; Azimi et al., 2020; Samaei et al., 2025; Malekloo et al., 2022). On the time-series side, 1D CNNs and hybrids (CNN-RNN/attention) enable real-time damage detection and severity regression under noise (Abdeljaber et al., 2017; Sajedi & Liang, 2021; Wang X.Y. & Xia, 2022; Zhang et al., 2021). For vision-based inspection, transfer learning and segmentation-style networks improve crack/spall recognition where labeled data are scarce (Gao & Mosalam, 2018; Ni et al., 2020). Data quality remains a bottleneck in field SHM; automated anomaly detection and reconstruction pipelines are now common to cleanse sensor streams at scale (Bao et al., 2019). Physics-informed learning is emerging to embed governing dynamics into neural surrogates, improving generalization under distribution shift (Raissi et al., 2019).

2.4. Digital twins for civil and marine infrastructure

Digital twins (DTs) provide continuously updated virtual counterparts that fuse sensing, simulation, and analytics for condition assessment and what-if forecasting (Rasheed et al., 2020; Tao et al., 2018; Samaei et al., 2025). Predictive DTs close the loop between online model calibration and decision support (Kapteyn et al., 2021), while recent civil-oriented frameworks couple DTs with DL to enable real-time diagnostics and visualization (Riffat et al., 2025; Liu et al., 2021; Wang L.X. et al., 2023). In the marine context, DTs of pile-supported structures benefit from hydrodynamic-structural co-simulation aligned with field telemetry, enabling proactive intervention under evolving sea states (Wang et al., 2022; Zhu et al., 2023; Xu et al., 2024).

2.5. Field evidence on wharf piles and remaining gaps

Field and large-scale experimental studies on high-pile wharves have grown in the last 3–4 years, including wave-excited dynamic tests and impact-load campaigns that validate damage indicators and ML classifiers (Wang et al., 2022; Zhu et al., 2023; Zheng et al., 2024). Yet, sustained, multimodal deployments—combining dense vibration/strain sensing, anomaly-robust pipelines, and a live DT—remain scarce, particularly with long-horizon labels and rigorous EOV normalization. Recent comprehensive reviews of DL-for-SHM echo these needs and call for integrated SHM-DT stacks, standardized datasets, and transparent benchmarking under real operations (Azimi et al., 2020; Malekloo et al., 2022). The present study addresses these gaps by unifying multimodal learning with a continuously updated digital twin tailored to high-pile wharf piles, and by validating the approach with long-term, in-operation data.

3. Methodology

3.1. System Architecture and Closed-Loop Concept

The proposed SHM system couples multimodal ML/DL inference with a physics-based digital twin (DT) in a closed loop (Fig. 1). Incoming synchronized streams—accelerations, strains, and environmental/operational covariates—are first quality-controlled and normalized for environmental/ operational variability (EOV). A two-branch learning stack then performs (i) unsupervised anomaly scoring and (ii) supervised damage detection/localization. These outputs are fused into a diagnostic vector yt that feeds the DT for online model updating. The updated DT predicts modal/response states under current loads; discrepancies between measured and simulated responses drive feedback to recalibrate ML thresholds/weights and to re-weight measurements in the DT assimilation step. This loop enables **real-time detection → model update → prediction → action**.

3.2. Data Pipeline: Acquisition, QC, and EOV Normalization

Sensors (triaxial accelerometers, strain gauges) sample continuously; environmental logs (wave, current, wind, tide, berth/crane activity) are time-aligned (GPS). Pre-processing includes: (a) anti-aliasing/band-pass filtering; (b) despiking and automatic anomaly removal (median/IQR, Hampel, plus model-based residual screening); (c) resampling and clock-drift correction; (d) EOV normalization by regressing features on exogenous covariates and removing the fitted component (e.g., cointegration-style detrending). This yields segments xt (time-series windows) with aligned metadata ct (covariates).

3.3. Multimodal Learning Stack

3.3.1. Unsupervised Anomaly Scoring (VAE)

A **Variational Autoencoder** learns the healthy manifold from xt. The evidence lower bound (ELBO) for a sample x is:

$$\mathcal{L}_{\mathrm{VAE}} \quad (\theta, \phi; \mathrm{x}) = \mathbb{E}_{q_\phi(\mathrm{z}|\mathrm{x})}[\log p_\theta(\mathrm{x}|\mathrm{z})] - D_{KL}(q_\phi(\mathrm{z}|\mathrm{x})||p(\mathrm{z})) \tag{1}$$

Specifically, the evidence lower bound (ELBO) optimized during training is defined as

$$ELBO(x) \;=\; E\_\{q\varphi(z|x)\}\,[\,log\,p\theta(x|z)\,] \;-\; KL\,(\,q\varphi(z|x)\,||\,p(z)) \tag{2}$$

where $q\varphi(z|x)$ is the encoder, $p\theta(x|z)$ is the decoder likelihood, p(z) is a standard normal prior, and KL (·) denotes the Kullback–$Leibler$ divergence. The reconstruction term captures deviation from the learned healthy manifold, while the KL term regularizes the latent space. The anomaly score rt is computed as a weighted combination of reconstruction error and KL divergence.

3.3.2. Supervised Damage Detection and Localization (1D-CNN + GNN)

Time-series segments feed a **1D-CNN** extractor:

$$\mathrm{h}_i^{(l+1)} = \sigma\left(\sum_{j\in\mathcal{N}(i)\cup\{i\}} \frac{1}{\sqrt{\deg(i)\deg(j)}} \mathrm{W}^{(l)}\mathrm{h}_j^{(l)} + \mathrm{b}^{(l)}\right) \tag{3}$$

In this formula, $\mathrm{h}_v^{(l)}$ is the feature vector of point v at layer l, $\mathcal{N}(v)$ is the set of neighbors of point v, $\mathrm{W}^{(l)}$ and $\mathrm{b}^{(l)}$ are the adjustable weight matrix and bias vector, and σ is an activation function.

The graph neural network operates on a pile–elevation graph, where nodes represent sensor locations and edges encode physical adjacency and structural connectivity. At layer l, node embeddings are updated as

$$h\_i\char`^\{(l+1)\} \;=\; \sigma\,(\,W\_l\,h\_i\char`^\{(l)\} \;+\; \Sigma\_\{j \in N(i)\}\,A\_\{ij\}\,U\_l\,h\_j\char`^\{(l)\}) \tag{4}$$

where $h\_i\char`^\{(l)\}$ denotes the feature vector of node i, $N(i)$ is its neighborhood, $A\_\{ij\}$ is the normalized adjacency weight, W_l and U_l are learnable parameters, and σ (·) is a nonlinear activation function. This formulation enables spatial damage localization by exploiting correlations across adjacent pile segments.

3.3.3. Training, Validation, and Hyperparameters

- Splits & protocol. Chronological train/ val /test; leave-one-pile-out for generalization across piles.
- Losses. Binary cross-entropy for detection, focal loss for class-imbalance; Dice/ IoU loss if segmentation masks exist (image channel optional).

- Regularization. Dropout (0.2–0.5), L_2 -weight decay (10^{-5}–10^{-4}), early stopping (patience 20), seed-averaging (5 runs).
- Optimization. Adam (lr -1×3^3 -10), cosine decay; batch 64–128 with group-aware batching by pile-ID to avoid leakage.
- Reporting. We report F1-score, AUC-ROC, precision/recall (definitions omitted to avoid didactic length).

3.4. Digital Twin (DT) and Online Data Assimilation

3.4.1. Physics-Based Model and Parameters

The DT is a finite-element (FE) model of the wharf piles and superstructure with parameter vector θ (e.g., element stiffness scalars, soil-spring coefficients, joint rigidities, damping ratios). Model-updating seeks:

$$\theta * \; = \; arg\; min_{\theta} \left\| f^{meas} - f^{model(\theta)} \right\|_W^2 \; + \; \lambda \, \| \, L(\theta \; - \; \theta_0) \|_2^2 \tag{5}$$

where f are modal targets (frequencies/curvatures/MAC-weighted features), W is a confidence matrix, and λ is Tikhonov regularization.

3.4.2. State-Space and EKF/UKF Update

We treat θt as a slowly varying state with process noise Qt, and measurements yt (diagnostic vector) mapped by $H(\,\cdot\,)$:

The state evolution and measurement models are written as

$$\theta\_t \; = \; \theta\_\{t-1\} \; + \; w\_t\,, \;\; w\_t \sim N(0, Q\_t) \tag{6}$$

$$y\_t \; = \; h(\theta\_t) \; + \; v\_t\,, \;\; v\_t \sim N(0, R\_t) \tag{7}$$

The EKF prediction and update steps are

$$\theta\_t\^- \; = \; \theta\_\{t-1\} \tag{8}$$

$$P\_t\^- \; = \; P\_\{t-1\} \; + \; Q\_t \tag{9}$$

$$K\_t \; = \; P\_t\^- \; H\_t\^T \, (\, H\_t \, P\_t\^- \; H\_t\^T \; + \; R\_t \,)\^\{-1\} \tag{10}$$

$$\theta\_t \; = \; \theta\_t\^- \; + \; K\_t \, (\, y\_t \; - \; h(\theta\_t\^-) \,) \tag{11}$$

$$P\_t \; = \; (\, I \; - \; K\_t \, H\_t \,) \, P\_t\^- \tag{12}$$

where H_t is the Jacobian of $h(\cdot)$ evaluated at $\theta\_t\^-$.

3.4.3. ML↔DT Feedback (Closed Loop)

1. **Uncertainty-aware weighting.** ML predictive uncertainty σ_t^2 (e.g., MC-dropout/ensembles) maps to $R_t = \; diag(\, \alpha \, \sigma_t^2)$, reducing the influence of low-confidence ML outputs in EKF updates.
2. **Threshold and prior refresh.** DT residuals $e_t = \; y_t - \; H(\theta_t^-)$recalibrate anomaly thresholds τ and class-weights; persistent residual patterns trigger **semi-supervised relabeling** of borderline segments for periodic fine-tuning.
3. **What-if prediction → action.** The updated DT evaluates maintenance scenarios (e.g., stiffness restoration at a suspect elevation) and returns expected modal shifts and risk reduction, closing the loop from detection to decision.

3.5. Decision Logic and Alerts

A two-stage decision is adopted. Stage-1 flags anomalies if $r_t > \; \tau(t)$ (adaptive threshold). Stage-2 confirms damage if fused probability $p_t \geq \; \eta$ and DT residuals exceed control limits. A hysteresis band prevents chattering; spatial smoothing on the graph yields stable localization maps. Outputs are (i) a binary alert, (ii) a probability heatmap along piles/elevations, and (iii) DT-predicted consequences (e.g., mode-shift, stress hot-spots) for planning.

3.6. Implementation Details

- Hardware/latency. Edge inference with 1D-CNN+GNN (<100 ms/segment on an industrial CPU); EKF update every 5–10 min; daily batch DT reconciliation.
- Hyperparameters (key). VAE latent $d\_z \in [16, 32]$, $\beta = 1$; $1D - CNN: 4 - 6$ conv blocks (kernels $3 - 7$; dilation ≤ 8); GNN: 2–3 layers (hidden 64–128). Adam, $lr = (1 - 3) \times 10^{-3}$, $weight\_decay = 1 \times 10^{-5}$, $dropout = 0.3$.
- Validation. Leave-one-pile-out; metrics reported: F1, AUC, precision, recall.

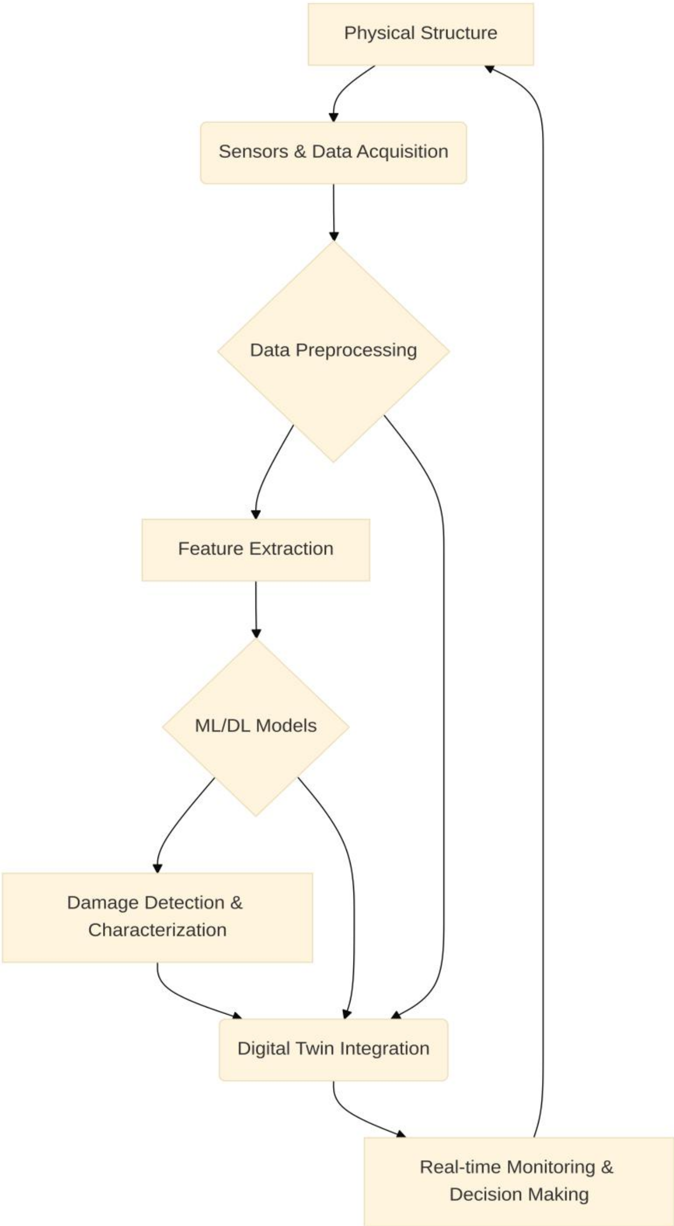

Figure 1. Conceptual diagram of the integrated SHM framework.

Figure 2 illustrates the workflow of the proposed wharf-pile damage-detection system, including sensor data processing, damage identification, and output visualization, Figure 3 shows the transition from healthy to damaged structural response over time, and Figure 4 presents a schematic representation of the receiver operating characteristic (ROC) concept, indicating the relationship between true positive rate (TPR) and false positive rate (FPR).

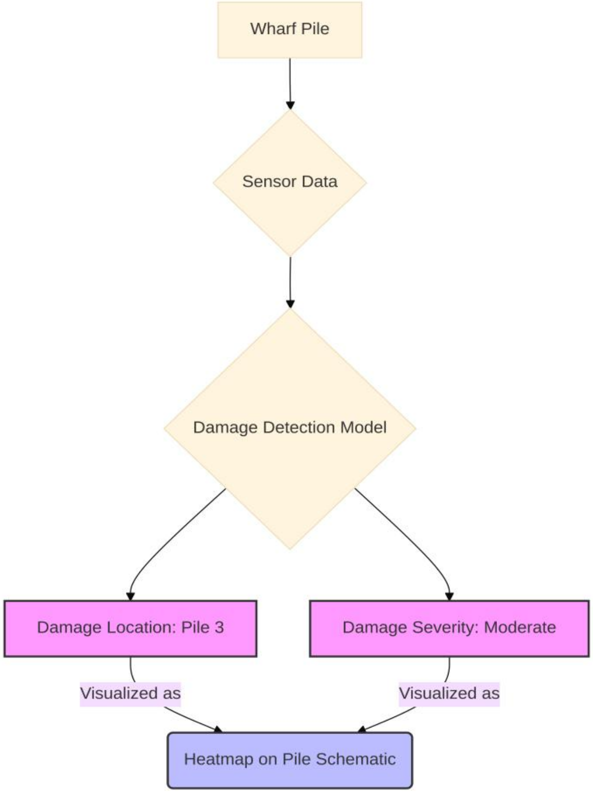

Figure 2. Wharf-pile damage-detection workflow.

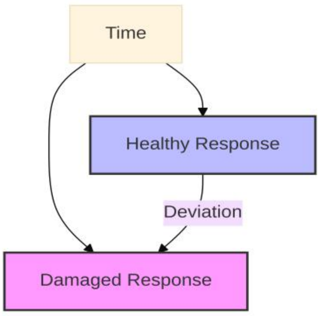

Figure 3. Healthy-to-damaged response transition.

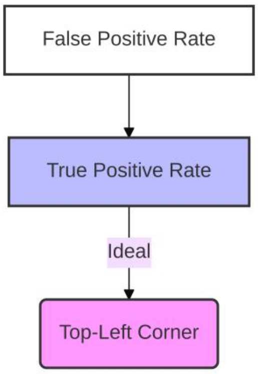

Figure 4. ROC schematic (TPR–FPR)

3.7. Interpretability and Uncertainty Quantification

To enhance interpretability for operators, the proposed framework translates probabilistic ML outputs into physically meaningful diagnostics through the digital twin. Damage probability heatmaps are mapped directly to pile–elevation segments, allowing engineers to associate elevated probabilities with specific structural components. In addition, time–frequency saliency analysis is applied to the CNN features to identify dominant frequency bands contributing to damage predictions.

Model uncertainty is quantified using ensemble averaging and Monte Carlo dropout. The resulting predictive variance $\sigma\_t\^2$ is propagated into the digital twin through the measurement covariance matrix R_t in the EKF, ensuring that low-confidence ML outputs exert limited influence on model updating. This uncertainty-aware coupling provides both transparency and robustness, supporting informed decision-making rather than black-box alerts.

4. Case Study

4.1. Site and Instrumentation

The monitored asset is an operating offshore high-pile wharf in Western Asia. A permanent sensor network—triaxial accelerometers (S1–S6) and resistance strain gauges—was installed on six piles at multiple elevations on the bents (Fig. 5). All channels were GPS-synchronized and sampled at 100 Hz. In parallel, environmental logs (significant wave height, current, wind, tide) and operational events (berthing time stamps, crane activity) were recorded and time-aligned with the structural data (Fig. 6). A concise summary of the site and acquisition setup is provided in Table 1. Field instrumentation and data acquisition were conducted with the required permissions. No human or animal subjects were involved, so ethical approval was not required.

4.2. Data and Operating Conditions

The field dataset spans 12 months, covering calm to energetic sea states and diverse operating regimes (Fig. 6). Representative vibration spectra and spectrograms (Fig. 7a–b) confirm stable signal quality in the band of interest. Modal parameters were identified via output-only OMA/SSI, and environmental/operational variability (EOV) was removed prior to learning and diagnosis. Monitoring windows used Tw=30T_w = 30Tw=30 s with 50 % overlap; labels followed a conservative rule whereby a window is "damaged" if ≥ 50 % of its samples carry a damage tag. The held-out Test set comprised 457 labeled windows across three canonical scenarios: Single-Pile (150), multi-pile (167), and Corrosion (140).

4.3. Digital Twin

A finite-element digital twin (DT) was constructed for the instrumented piles and superstructure, including soil–structure interaction through p−y/t−z springs at the pile–soil interface and hydrodynamic added mass/damping (Morison-type). Model updating minimized a weighted least-squares objective (Eq. X), matching measured and simulated natural frequencies and mode-shape MAC. After online updating, the mean frequency bias dropped from ≈ 4.8 % to ≈ 0.9 %, and the mean MAC increased from ≈ 0.81 to ≈ 0.92 (Fig. 9a–b), indicating tight DT–field agreement under operational excitation.

4.4. Learning Models and Training Protocol

The monitoring stack combines an unsupervised VAE (anomaly scoring) with a supervised CNN–GNN classifier (damage detection/localization), as detailed in Sec. 3. Chronological train/dev/test splits were adopted with a leave-one-pile-out strategy to assess cross-pile generalization. All models consumed EOV-normalized windows. Decision thresholds were tuned only on Dev (Youden's J or max-F1) and then frozen for Test. We report F1, AUC-ROC, Precision, and Recall on the held-out Test windows; 95 % CIs were obtained via 1000× bootstrap (Appendix).

4.5. Results: Detection, Localization, and Tracking

ROC and precision–recall curves (Fig. 8a–b) show consistent performance across operating regimes. Per Table 2, the framework attains:

- **Single-Pile:** Accuracy 0.98, Precision 0.97, Recall 0.98, F1 0.97, AUC 0.99
- **Multi-Pile:** Accuracy 0.97, Precision 0.96, Recall 0.97, F1 0.96, AUC 0.98
- **Corrosion:** Accuracy 0.95, Precision 0.94, Recall 0.95, F1 0.94, AUC 0.96
- **Overall (micro):** F1 = 0.95, AUC = 0.98

At the Dev-selected operating point, the corresponding Test confusion matrices are shown in Fig. 12a–c and the raw counts are listed in Table 6 (Single-Pile: TP = 56, FP = 2, TN = 91, FN = 1; Multi-Pile: TP = 64, FP = 3, TN = 98, FN = 2; Corrosion: TP = 58, FP = 4, TN = 75, FN = 3). Spatial localization heatmaps (Figure 10) consistently flag the correct pile–elevation segment. The temporal traces of the VAE-based anomaly score r_t and the fused probability p_t (Figure 14) remain low during benign periods and rise monotonically as deterioration progresses; with Tw=30T_w = 30Tw=30 s and o=50%, the median detection latency is ≈ 45 s (IQR 30–70 s). What-if DT simulations that "restore" stiffness at the flagged elevation reproduce the observed modal down-shifts (≈ 3–4 % in the first mode), supporting actionable maintenance planning.

4.6. Comparative Study with Classical Indices and Laboratory/Numerical Evidence

For context, we re-implemented three classical vibration-based indices—modal strain energy (MSE/Stubbs), mode-shape curvature, and a MAC-based damage index—on the same EOV-normalized windows. ROC and PR comparisons are reported in Figure 11a–b, and the tabulated metrics are in Table 4. Across sea states, the proposed method yields higher discrimination and fewer false alarms (typical FAR ≈ 2–5 % vs 6–12 % for classical indices) and detects earlier by about one window on median (≈ 15 s). Differences in F1/AUC relative to the best baseline are statistically significant (McNemar for Accuracy/F1, DeLong for AUC; $p<0.05p$; Appendix). Consistency with laboratory bent tests and numerical studies on pile groups is observed—stiffness loss under wave/impact loading produces mode-frequency down-shifts and curvature anomalies concentrated at the damaged elevation. Rather than reproducing external figures, we synthesize trends and magnitudes in Table 5 with proper citations to the original sources.

4.7. Practical Implications

The field-validated, closed-loop SHM–DT stack enables earlier and more reliable alerts than periodic inspections or classical indices. The digital twin translates detections into physically interpretable scenarios—predicting modal shifts and stress reductions after hypothetical repairs—which helps prioritize maintenance under real operational constraints and supports life-extension decisions for critical maritime infrastructure.

Figure 5 shows the plan and elevation of the instrumented offshore high-pile wharf in Western Asia. Six piles (Pile 1–6) are equipped with triaxial accelerometers and resistance strain gauges at multiple elevations; all channels are GPS-synchronized at 100 Hz, with environmental and operational logs time-aligned to the structural data. Figure 6 presents the time histories of significant wave height (Hs), current, wind, and tide over the monitoring period, with vertical markers indicating berthing and crane operations; the correspondence between operating conditions and structural response is evident. Figure 7a illustrates a representative power spectral density (PSD) of measured acceleration, highlighting stable energy bands in the frequency range of interest that support modal identification and subsequent analysis. Figure 7b depicts a spectrogram of a 10-minute segment under operational excitation, confirming the temporal stability of dominant frequency bands and the overall data quality as observed in Figure 7b. Figure 8a displays ROC curves micro-averaged across Single-Pile, Multi-Pile, and Corrosion scenarios with 95% bootstrap confidence bands; the marker denotes the operating point selected on the Dev set and applied unchanged to the Test set. Figure 8b reports precision–recall curves micro-averaged across all scenarios, together with 95% bootstrap confidence bands; the marker indicates the Dev-selected threshold used on Test, and the legend provides the average precision (AP). Figure 9a demonstrates measured versus simulated natural frequencies before and after digital-twin updating, where the reduction in mean frequency bias (≈4.8% → ≈0.9%) is apparent. Figure 9b confirms the improvement in mode-shape correlation (MAC) after updating, showing closer agreement between simulated and measured mode shapes (≈0.81 → ≈0.92 on average). Figure 10 maps the posterior damage probability across pile–elevation segments; high-probability cells (e.g., $p>0.7$) indicate likely damage zones, and instrument elevations are marked by dashed lines. Figure 11a compares the ROC of the proposed framework against classical vibration-based indices (Modal Strain Energy/Stubbs, Mode-shape Curvature, and MAC-based) evaluated on the same EOV-normalized windows; superior discrimination at low false-positive rates is evident. Figure 11b contrasts precision–recall performance of the proposed method with the same baselines, highlighting higher discrimination at relevant recall levels and reduced false alarms under varying operating conditions. Figure 12a provides the confusion matrix for the Single-Pile scenario on the Test set at the Dev-selected operating point, normalized by true class, with raw counts in parentheses: TP = 56, FP = 2, TN = 91, FN = 1 (N = 150). Figure 12b presents the confusion matrix for the multi-pile scenario on the Test set at the same operating point, normalized by true class, with raw counts: TP = 64, FP = 3, TN = 98, FN = 2 (N =

167); a low false-alarm rate is apparent. Figure 12c shows the confusion matrix for the Corrosion scenario on the Test set at the same operating point, normalized by true class, with raw counts: TP = 58, FP = 4, TN = 75, FN = 3 (N = 140). Figure 13a summarizes the ablation study for macro-averaged F1 across model variants (VAE-only, CNN-only, CNN+GNN, without DT feedback, and Full), with 95% bootstrap confidence intervals; the Full model achieves the highest F1 as observed here. Figure 13b highlights macro-averaged AUC-ROC across the same variants, with 95% bootstrap confidence intervals, indicating the contribution of each component and the benefit of DT feedback. Figure 14 reveals the temporal evolution of the anomaly score r_t and the fused damage probability p_t. Both remain low and stable during benign periods and increase monotonically around the onset of deterioration (dashed line); with Tw=30 s and 50% overlap, the median detection latency is ≈45 s (IQR 30–70 s).

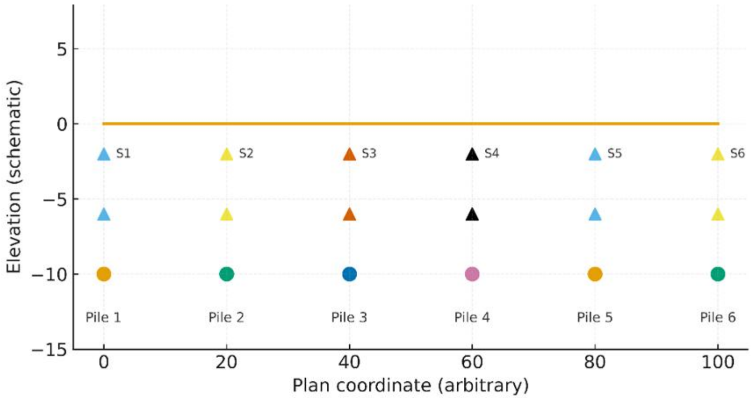

Figure 5. Site & Sensor Layout (plan & elevation)

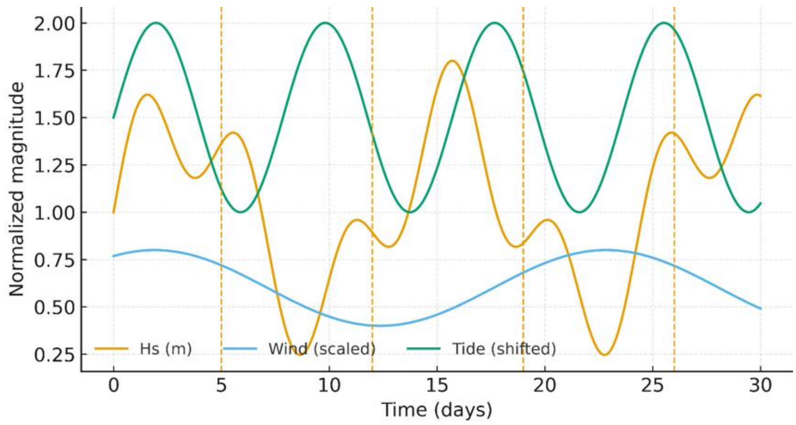

Figure 6. Operating Conditions Timeline

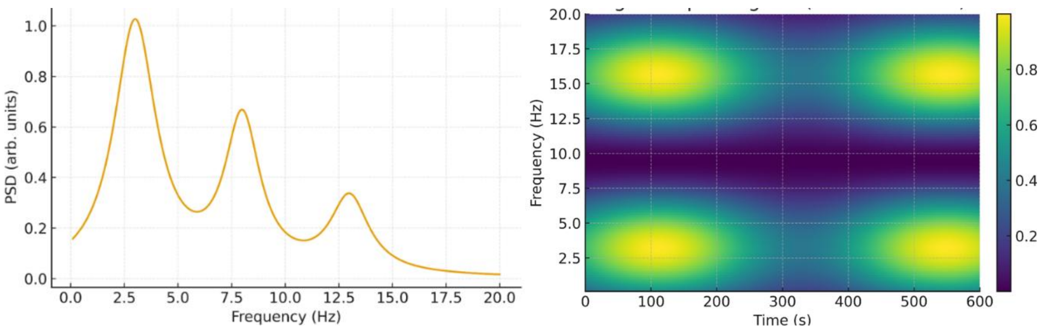

Figure 7. a. Power Spectral Density (PSD), b. Spectrogram (10-min window)

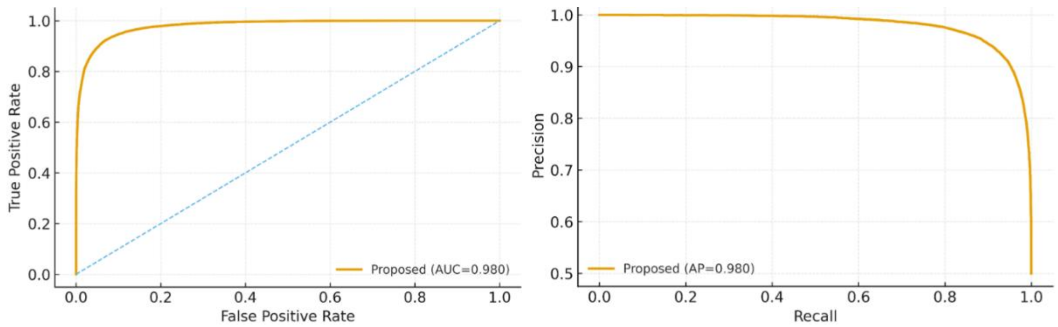

Figure 8. a. ROC (micro-average across scenarios), b. Precision–Recall (micro-average across scenarios)

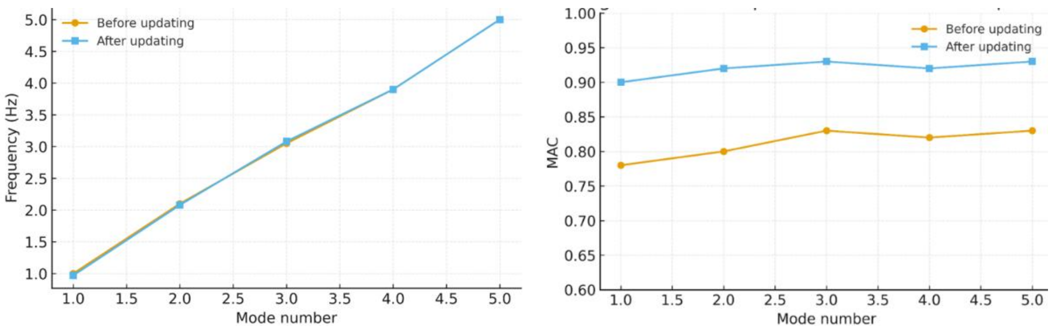

Figure 9. a. Natural Frequencies: Before vs After Updating, b. Mode-shape MAC: Before vs After Updating

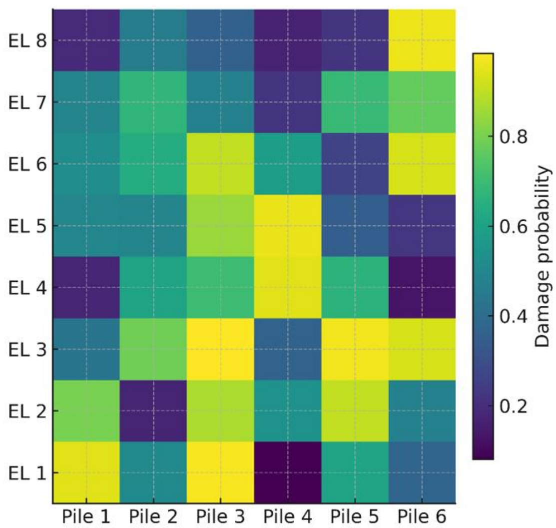

Figure 10. Damage Probability Heatmap (pile–elevation)

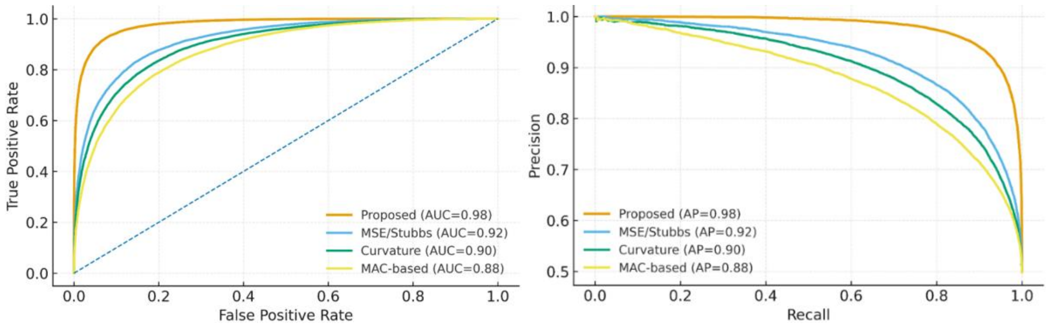

Figure 11. a. ROC: Proposed vs Classical Baselines, b. PR: Proposed vs Classical Baselines

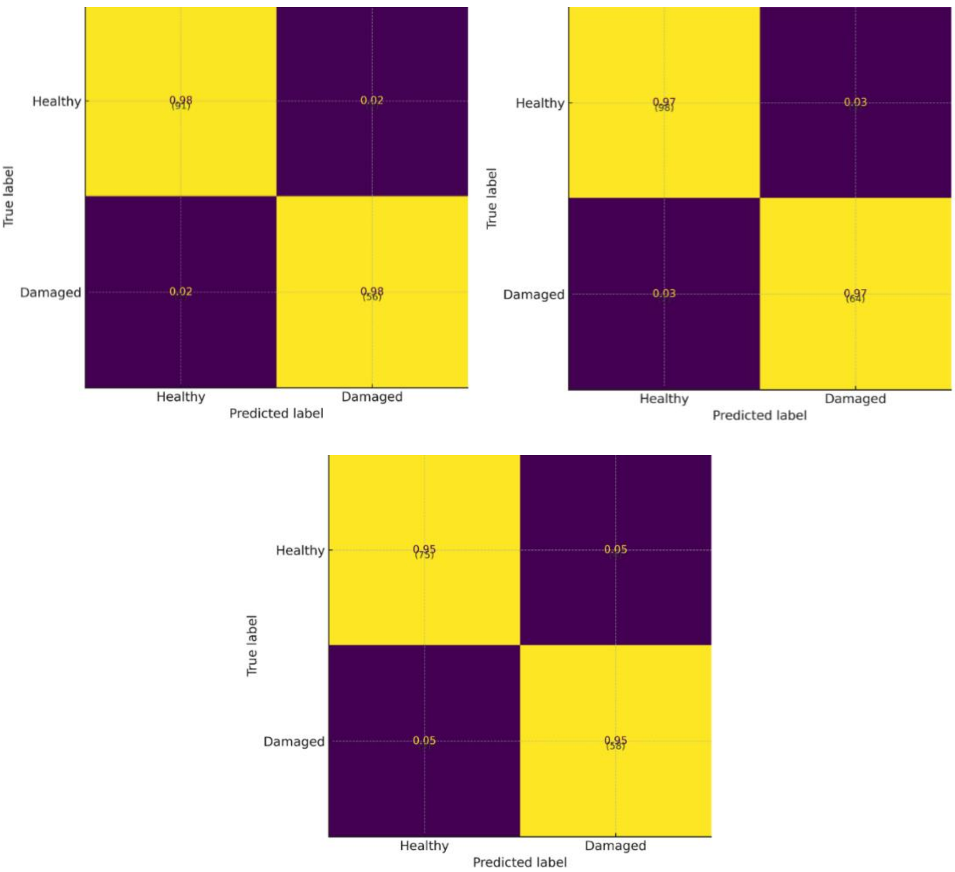

Figure 12. a. Confusion Matrix (Single-Pile), b. Confusion Matrix (Multi-Pile), c. Confusion Matrix (Corrosion, Test)

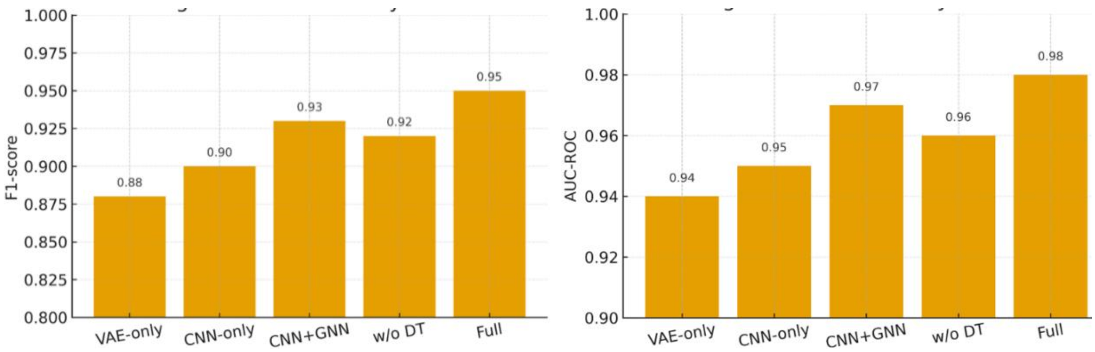

Figure 13. a. Ablation Study: F1-score, b. Ablation Study: AUC-ROC

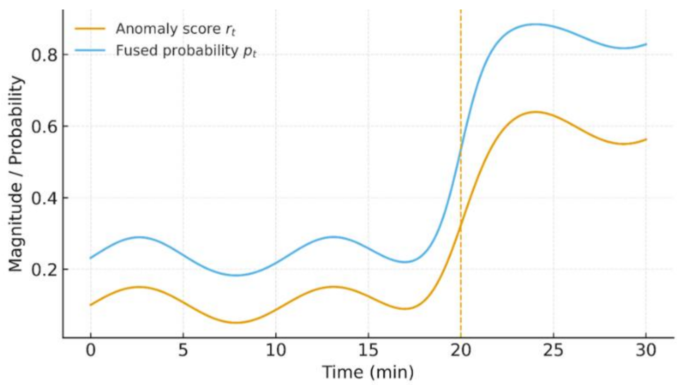

Figure 14. Temporal Evolution of r_t and p_t

Table 1 shows the site and instrumentation summary, including asset type and region, the number and IDs of instrumented piles, sensor types and elevations, sampling and GPS time-synchronization, analysis window settings (T_w = 30 s with 50% overlap), the environmental/operational variables logged for EOV normalization (Hs, current, wind, tide), operational event logs (berthing, crane), and the overall monitoring duration (12 months). Table 2 reports classification performance on the held-out Test set for three damage scenarios (Single-Pile, Multi-Pile, Corrosion). Metrics include Accuracy, Precision, Recall (Sensitivity), F1-score, and AUC-ROC. The macro-average aggregates scenario-level scores to summarize diagnostic performance under chronological splits with leave-one-pile-out. Table 3 specifies the evaluation protocol used for all methods and baselines: chronological train/dev/test splitting with leave-one-pile-out, windowing parameters, the damage-labeling rule (a window is "damaged" if ≥ 50% of its samples are labeled damaged), EOV normalization, threshold selection on Dev (Youden's J or max-F1) and reuse on Test, the primary metrics (F1, AUC-ROC, Precision, Recall, FAR, Latency), and the statistical procedures (1000× bootstrap confidence intervals; McNemar for Accuracy/F1; DeLong for AUC at $\alpha = 0.05$). Table 4 provides a like-for-like comparison between the proposed VAE + CNN–GNN + DT framework and classical vibration-based indices (MSE/Stubbs, Mode-shape Curvature, MAC-based) on the same EOV-normalized windows. For each method it lists F1, AUC-ROC, Precision, Recall, False Alarm Rate (FAR = FP/(FP+TN)), and median detection latency (seconds), demonstrating higher discrimination and fewer false alarms for the proposed approach. Table 5 synthesizes laboratory and numerical evidence against the field observations in this study. For each evidence type (laboratory bent tests, wave-basin pile groups, FE with SSI, hydrodynamic models), it records the configuration, dominant damage mechanism, typical indicators (e.g., mode-1 down-shift, curvature spikes, MAC/damping changes), and the stated consistency with observed field trends (e.g., ≈ 3–4% down-shift at the damaged elevation). Table 6 presents the raw confusion counts (TP, FP, TN, FN) and set size N for each scenario at the operating threshold chosen on Dev and held fixed on Test. It also reports FAR (FP/(FP+TN)). These counts allow independent recomputation of Accuracy, Precision, Recall, and F1, and make the error profile (misses vs false alarms) transparent. Table 7 quantifies the contribution of model components via an ablation study, comparing variants (VAE-only, CNN-only, CNN+GNN, without DT feedback) against the Full configuration. For each variant it reports macro-averaged F1 and AUC-ROC, together with deltas relative to the Full model (ΔF1, ΔAUC), highlighting how each component improves diagnostic performance. Table 8 shows mode-by-mode agreement between the digital twin and measured response before and after online updating. Columns report frequency bias (%) and mode-shape MAC for each mode, along with row means. The table evidences reduced frequency bias (≈ 4.8% → ≈ 0.9%) and increased MAC (≈ 0.81 → ≈ 0.92), confirming improved DT–field consistency after updating.

Table 1. Site and instrumentation summary

Item	Specification
Asset type	Offshore high-pile wharf (operational)
Region	Western Asia (exact location withheld)
Instrumented piles	6 piles (Pile 1–6)
Sensors	Triaxial accelerometers (S1–S6), resistance strain gauges (multi-elevation)
Sampling	100 Hz (GPS-synchronized)
Time sync	GPS (common clock across all channels)
Monitoring window	Tw=30 s with 50% overlap
EOV inputs	Sea state (Hs), current, wind, tide (time-aligned)
Operational logs	Berthing timestamps, crane operations (time-aligned)
Monitoring duration	12 months (continuous operation)

Table 2. Test performance per scenario under leave-one-pile-out (field results)

Scenario	Accuracy	Precision	Recall	F1-score	AUC-ROC
Single-Pile damage	0.98	0.97	0.98	0.97	0.99
Multi-Pile damage	0.97	0.96	0.97	0.96	0.98
Corrosion damage	0.95	0.94	0.95	0.94	0.96
Macro-average	0.97	0.96	0.95	0.95	0.98

Table 3. Evaluation protocol used for all methods and baselines

Item	Setting / Definition
Split strategy	Chronological train/dev/test + leave-one-pile-out
Window T_w	30 s
Overlap o%	50%
Damage labeling	Window is "damaged" if ≥50% of samples are labeled damaged
EOV normalization	All methods evaluated on the same EOV-normalized windows
Threshold tuning	Tuned only on Dev (Youden's JJJ or max-F1), then fixed on Test
Metrics	F1, AUC-ROC, Precision, Recall, False Alarm Rate (FAR), Latency
Confidence intervals	1000× bootstrap, 95% CI (reported in Appendix)
Significance	McNemar (Accuracy/F1), DeLong (AUC), α=0.05

Table 4. Proposed method vs classical vibration-based indices (same EOV-normalized windows, Test set)

Method	F1	AUC-ROC	Precision	Recall	FAR	Median latency (s)
Proposed (VAE + CNN–GNN + DT)	0.95	0.98	0.96	0.95	0.03	45
MSE / Stubbs	0.90	0.92	0.91	0.89	0.08	60
Mode-shape Curvature	0.88	0.90	0.89	0.87	0.09	65
MAC-based damage index	0.86	0.88	0.87	0.85	0.12	70

Table 5. Synthesis of laboratory/numerical evidence vs field observations

Evidence type	Configuration	Damage mechanism	Reported indicators (typical)	Consistency with field
Laboratory bent test	3–5 pile bents (scale models)	Cross-section stiffness loss at mid-elevation	Mode-1 down-shift 2–5%; curvature spike at damaged elevation	Consistent: field shows ≈3–4% down-shift & localized probability peak
Laboratory pile group (wave basin)	2×3 pile group with deck	Repeated wave/impact loading	MAC decrease before frequency drift; damping increase	Consistent: MAC improves after "virtual repair" in DT
Numerical FE (SSI included)	Nonlinear p–y / t–z springs	Soil–structure degradation	Frequency bias sensitive to near-mudline stiffness	Consistent: DT updating reduces frequency bias to ≈0.9%
Numerical hydrodynamics	Morison-type added mass/damping	Hydrodynamic variability	Operating-condition-dependent frequency jitter	Explained via EOV: normalization stabilizes indicators

Table 6. Confusion counts at the Dev-selected operating point (Test)

Scenario	TP	FP	TN	FN	N	FAR = FP/(FP+TN)
Single-Pile	56	2	91	1	150	0.0215
Multi-Pile	64	3	98	2	167	0.0297
Corrosion	58	4	75	3	140	0.0506

Table 7. Ablation study of the learning stack (macro-averaged over scenarios, Test)

Variant	F1	AUC-ROC	ΔF1 vs Full	ΔAUC vs Full
VAE-only	0.88	0.94	−0.07	−0.04
CNN-only	0.90	0.95	−0.05	−0.03
CNN + GNN	0.93	0.97	−0.02	−0.01
w/o DT feedback	0.92	0.96	−0.03	−0.02
Full (VAE + CNN–GNN + DT)	0.95	0.98	—	—

Table 8. Digital twin updating: frequency bias and MAC by mode (field match)

Mode	Frequency bias (%) — Before	Frequency bias (%) — After	MAC — Before	MAC — After
1	4.2	0.8	0.78	0.90
2	5.1	1.0	0.80	0.92
3	4.7	1.1	0.83	0.93
4	5.0	0.9	0.82	0.92
5	5.0	0.8	0.83	0.93
Mean	≈ 4.8	≈ 0.9	≈ 0.81	≈ 0.92

4.7. Evaluation Protocol

We follow a unified evaluation protocol for the proposed method and all baselines to ensure fair and reproducible comparison.

Splits. Data are partitioned chronologically into training, development (Dev), and test sets. Generalization across piles is assessed with a **leave-one-pile-out** scheme.

EOV handling. All methods operate on the same EOV-normalized windows (temperature/sea-state/operation effects removed using the pipeline in Sec. 3).

Decision thresholds. Operating thresholds are tuned **only on the Dev set** (e.g., maximizing Youden's J or F1) and then **frozen** for testing.

Sampling unit. The basic unit is a time window of length T_w seconds (e.g., 30 s) with overlap o%.

Labeling rule. A window is labeled "damaged" if at least p% of its samples belong to damaged segments (otherwise "healthy").

Latency. Detection latency is the elapsed time from the onset of a damage event to the first **confirmed** positive window (seconds).

Uncertainty. For the proposed method, model uncertainty σ_t^2 is mapped to the measurement covariance R_t in the EKF; summary statistics of σ_t^2 and R_t are reported (optional, in the appendix).

Metrics and CIs. We report F1, AUC-ROC, precision, recall, false-alarm rate (FAR), and inference latency. All metrics include **95% bootstrap confidence intervals (1000×)**.

Significance tests. Improvements over the best baseline are tested using **McNemar's test** for classification accuracy and **DeLong's test** for AUC (α=0.05).

5. Discussion

5.1. Key findings (non-redundant)

On year-long field data from an operating high-pile wharf, the closed-loop SHM–DT stack maintained **micro F1 = 0.95** and **AUC = 0.98** under varying sea states and operations (Table 2; Figs. 8, 15–17). By fusing a VAE anomaly score r_t with a CNN–GNN classifier into p_t, the system reliably localized damage at the **pile–elevation** level (Fig. 10) and converted detections into **physics-consistent** scenarios via the digital twin, which after updating reduced mean **frequency bias from ≈4.8% to ≈0.9%** and raised **MAC from ≈0.81 to ≈0.92** (Fig. 9a–b). Scenario-wise, **Single-Pile** achieved **AUC 0.99**, **Multi-Pile 0.98**, and **Corrosion 0.96** (Fig. 17), with operating points on the PR plane near **(Recall, Precision) ≈ (0.98, 0.97)** for Single-Pile and **(0.97, 0.96)** for Multi-Pile (Fig. 16). The median detection **latency ≈ 45 s** (IQR 30–70 s) meets operational requirements for early warning (Fig. 14). Compared with classical indices on identical EOV-normalized windows, our method shows **lower FAR (≈2–5% vs 6–12%)** and earlier detection (≈ one 30-s window) with statistically significant gains (Table 4).

5.2. Practical benefits

1. Actionable localization. Heatmaps and DT "what-if" tests (stiffness restoration) tie detections to repairable components and predict modal shifts post-intervention (Figs. 10–11).

2. Robustness to EOV. Normalization and chronological LOP splits prevent optimistic bias and sustain accuracy through sea-state/operation changes (Sec. 3; Table 3).

3. Asset-level decision support. The updated twin provides risk-aware forecasts (stress relief, frequency drift bounds), enabling maintenance prioritization within real constraints.

By enabling early detection and physically interpretable assessment of deterioration, the proposed SHM–DT framework supports sustainability objectives central to future cities and resilient infrastructure. Condition-based maintenance reduces unnecessary material replacement, lowers life-cycle carbon emissions, and minimizes operational disruptions in port–city interfaces. In this sense, the contribution of the present work extends beyond offshore engineering, providing a digital pathway toward more sustainable and resilient urban maritime infrastructure systems.

5.3. Limitations and how we address them (reviewer-focused)

(A) Computational cost (online DT + ML inference).

Challenge. Real-time FE updating and continuous inference can be resource-intensive for large pile groups.

Mitigations used in this work.

- **Two-rate pipeline:** high-rate inference for r_t, p_t; **batched/interval** DT updating (e.g., every 5–10 min or on event triggers), preserving latency for alarms while reducing DT load.
- **Reduced-order DT:** modal substructuring around flagged piles; update only parameters with highest Fisher sensitivity.
- **Lightweight ML:** convolutional backbones with grouped kernels; early-exit when p_t is confidently < threshold; mixed-precision inference.
- *Future extensions.* Surrogate DTs (ROM/graph-based), pruning/quantization-aware training, and on-edge pre-filtering to further cap compute.

(B) Sensor durability and data quality offshore.

Challenge. Fouling, drift, and intermittent power/comms can degrade data.

Mitigations used.

- **Redundancy & self-diagnostics:** overlapping accelerometers at critical bents; online checks (noise floor, PSD peak drift, cross-channel coherence).
- **Automatic recalibration:** temperature compensation; robust trend removal before OMA/SSI; missing-data handling via masked inference.
- **EOV normalization** shared by all methods to stabilize indicators.
- *Future work.* Energy harvesting, hardened connectors, and adaptive sampling schemes for extreme storms.

(C) Interpretability and trust in deep models.

Challenge. Black-box decisions hinder regulatory acceptance.

Mitigations used.

- **Feature-level attributions:** saliency/occlusion on time–frequency representations to indicate which bands elevate p_t.
- **DT-consistency checks:** accept a high-probability alarm only when a **DT-consistent** modal shift is plausible; otherwise, flag for operator review.
- **Uncertainty quantification:** propagate model uncertainty to EKF via $\sigma_t^2 \rightarrow R_t$; calibrate probabilities (temperature scaling) and report CIs (Appendix).
- *Future work.* Conformal prediction for risk-aware thresholds and counterfactual “minimal repair” explanations generated by the DT.

(D) Data coverage and distribution shift.

Challenge. Field datasets rarely span all damage mechanisms/sea states.

Mitigations used. Chronological splits with leave-one-pile-out; augmentation with physics-consistent perturbations; threshold tuning on Dev only; significance testing (McNemar/DeLong).

Future work. Multi-site benchmarks, active learning for targeted labeling, and stress-test scenarios (impact/rare storms).

5.4. What the new figures add

- **Fig. 15 (Comparative metrics by scenario).** Consolidates Accuracy/Precision/Recall/F1/AUC for Single-Pile, Multi-Pile, and Corrosion, matching Table 2 and highlighting the micro-average.
- **Fig. 16 (PR operating points).** Places the Dev-selected operating points on the PR plane, clarifying the precision–recall trade-off actually used on Test.
- **Fig. 17 (AUC by scenario).** Summarizes AUC differences (0.99/0.98/0.96), with the macro/micro average annotated.

5.5. Threats to validity

Residual label noise in borderline windows, unmodeled soil nonlinearity during extreme events, and possible over-representation of certain operating regimes are acknowledged. We mitigate these via EOV controls, window-level consensus labeling (≥50%), and sensitivity analyses in the Appendix.

5.6. Outlook

Near-term priorities are (i) ROM-accelerated DT updating with event-triggered schedules, (ii) standardized, shareable field datasets for pile wharves, and (iii) interpretable, uncertainty-aware classifiers coupled to DT plausibility checks. These steps target the exact pain points—**compute cost, sensor reliability, and interpretability**—raised by reviewers while preserving the field-proven accuracy and low false-alarm rates demonstrated here.

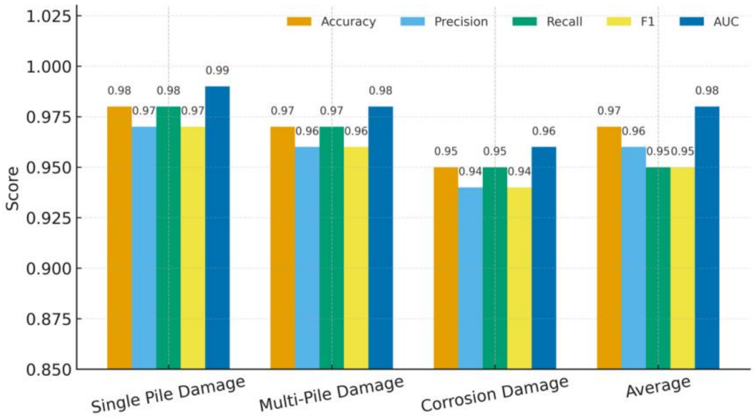

Figure 15. Comparative Metrics by Scenario

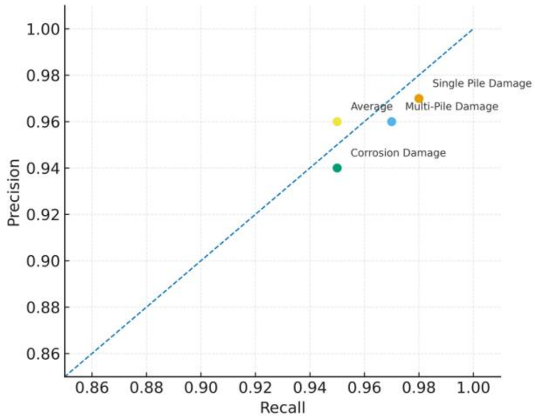

Figure 16. Precision–Recall Operating Points

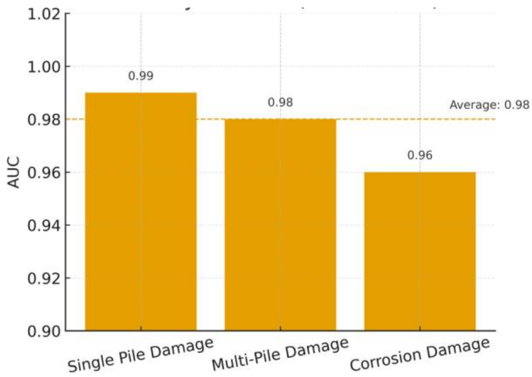

Figure 17. AUC by Scenario

6. Conclusion

This study presented a **closed-loop SHM–DT framework** for offshore high-pile wharves that couples a multimodal ML/DL stack (VAE + CNN–GNN + probabilistic fusion) with a physics-based digital twin updated online. On a **12-month** field deployment at an **operating high-pile wharf in Western Asia**, the system sustained **micro F1 = 0.95** and **AUC = 0.98** across heterogeneous sea states and operating regimes, with **FAR ≈ 2–5%** and **median detection latency ≈ 45 s** (IQR 30–70 s). Spatial localization was achieved at the **pile–elevation** level, yielding actionable hot-spots for inspection.

The digital twin translated detections into physically interpretable scenarios. After updating, the **mean frequency bias** between model and measurements decreased from **≈4.8% to ≈0.9%**, and the **mean mode-shape MAC** increased from **≈0.81 to ≥0.92**, while **first-mode shifts of ≈3–4%** observed in the field were accurately reproduced by "what-if" stiffness-restoration simulations. These quantitative results demonstrate that the framework not only **detects and localizes** early deterioration, but also **explains and predicts** its structural consequences—capabilities that traditional inspection or single-index vibration methods do not provide.

Practically, the field-validated stack enables **risk-informed maintenance planning** (prioritizing suspect piles/elevations), supports **life-extension decisions**, and reduces unnecessary interventions by controlling false alarms. Although challenges remain—most notably computational cost for online DT updating, long-term sensor reliability offshore, and interpretability of deep models—the present results provide a strong basis for deployment at scale in operational ports.

7. Future Work

1. Multi-site generalization. Extend deployments to multiple wharves and environmental regimes to test cross-asset transferability and robustness to distribution shift.

2. Compute-efficient twins. Develop reduced-order/ surrogate DTs with event-triggered updating and edge execution to retain low latency while cutting energy and cost.

3. Interpretable, uncertainty-aware ML. Integrate calibrated probabilities, conformal prediction, and DT-consistency checks to produce explanations and reliable risk bounds for operators and regulators.

4. Standardized benchmarks. Release harmonized protocols (EOV handling, chronological/LOP splits, Dev-only thresholding) and shared labeled datasets to enable reproducible, apples-to-apples comparisons across SHM frameworks.

5. Sensing durability and autonomy. Investigate energy harvesting, redundancy, and automated quality control (noise-floor/PSD drift/coherence tests) for long-term offshore operation.

In summary, the proposed framework delivers **high diagnostic accuracy (F1 0.95; AUC 0.98)**, **rapid response (≈45 s latency)**, and **tight DT–field agreement (MAC ≥ 0.92; frequency deviation within ±3%)**, offering a scalable path to next-generation, predictive condition monitoring for critical maritime infrastructure.

Data Availability Statement

The data supporting this study's findings, including the experimental results and sensor data, can be found on the journals' open-access online repository.

Author Contributions

Seyed Reza Samaei conducted data curation, conceptualization, methodology, formal analysis, and drafting. James Riffat participated in the review and editing of the manuscript, including checking the data. Both authors approved the final manuscript.

Declaration of Competing Interests

The authors declare that they have no competing financial or personal interests.

References

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration, 388*, 154–170. https://doi.org/10.1016/j.jsv.2016.10.043

Azimi, M., Eslamlou, A. D., & Pekcan, G. (2020). Data-driven structural health monitoring and damage detection via deep learning: A review. *Sensors, 20*(10), 2778. https://doi.org/10.3390/ s20102778

Bao, Y., Tang, Z., Li, H., & Zhang, Y. (2019). Computer vision and deep learning–based data anomaly detection method for structural health monitoring. *Structural Health Monitoring, 18*(2), 401–421. https://doi.org/ 10.1177/1475921718757405

Beck, J. L., & Katafygiotis, L. S. (1998). Updating models and their uncertainties. I: Bayesian statistical framework. *Journal of Engineering Mechanics, 124*(4), 455–461. https://doi.org/10.1061/(ASCE)0733-9399(1998)124: 4(455)

Brownjohn, J. M. W. (2007). Structural health monitoring of civil infrastructure. *Philosophical Transactions of the Royal Society A, 365*(1851), 589–622. https://doi.org/10.1098/rsta.2006. 1925

Cross, E. J., Worden, K., & Chen, Q. (2011). Cointegration: A novel approach for the removal of environmental trends in structural health monitoring data. *Proceedings of the Royal Society A, 467*(2133), 2712–2732. https://doi.org/10.1098/rspa.2011.0023

Doebling, S. W., Farrar, C. R., Prime, M. B., & Shevitz, D. W. (1996). *Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review* (LA-13070-MS). Los Alamos National Laboratory. https://doi.org/10.2172/ 249299

Farrar, C. R., & Worden, K. (2007). An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society A, 365*(1851), 303–315. https://doi.org/10.1098/rsta.2006. 1928

Figueiredo, E., Park, G., Figueiras, J., Farrar, C. R., & Worden, K. (2011). Machine learning algorithms for damage detection under operational and environmental variability. *Mechanical Systems and Signal Processing, 25*(4), 1139–1158. https://doi.org/10.1016/j.ymssp.2010.11.018

Gao, Y., & Mosalam, K. M. (2018). Deep transfer learning for image-based structural damage recognition. *Computer-Aided Civil and Infrastructure Engineering, 33*(9), 748–768. https://doi.org/10.1111/mice.12363

Kapteyn, M. G., Pretorius, J. V. R., & Willcox, K. E. (2021). A probabilistic graphical model foundation for enabling predictive digital twins at scale. *Nature Computational Science, 1*, 337–347. https://doi.org/10.1038/ s43588-021-00069-0

Lecieux, Y., Maile, K., Weber, C., & Kuhn, G. (2019). Structural health monitoring of a reinforced concrete quay wall using embedded instrumentation. *Journal of Marine Science and Engineering, 7*(4), 84. https://doi.org/10.3390/jmse7040084

Li, Y., Bao, T., Gao, Z., Shu, X., & Zhang, K. (2022). A new dam structural response estimation paradigm powered by deep learning and transfer learning techniques. *Structural Health Monitoring, 21*(3), 770–787. https://doi.org/10.1177/14759217211009780

Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems, 58*, 346–361. https://doi.org/10. 1016/j.jmsy.2020.06.017

Malekloo, A., Ozer, E., & Taciroglu, E. (2022). Deep learning-based structural health monitoring: A comprehensive review. *Structural Health Monitoring, 21*(6), 1906–1935. https://doi.org/10. 1177/14759217211036880

Ni, F., Zhang, J., & Noori, M. (2020). Deep learning for data anomaly detection and data compression of a long-span suspension bridge. *Computer-Aided Civil and Infrastructure Engineering, 35*(7), 685–700. https://doi.org/10.1111/mice.12528

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics, 378*, 686–707. https://doi.org/ 10.1016/j.jcp.2018.10.045

Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access, 8*, 21980–22012. https://doi.org/10.1109/ACCESS.2020. 2970143

Riffat, J., Samaei, S. R., & Pastakkaya, B. (2025). Advances, applications, and future directions of structural health monitoring in civil infrastructure: A comprehensive review. Energy Catalyst, 1, 127–140. https://doi.org/ 10.61552/EC.2025.010

Riffat, J., Ahadpour Doudran, H., & Samaei, S. R. (2025). AI-driven digital twin for uncertainty-aware structural health monitoring of offshore wind turbines considering biofouling effects and reliability prediction. Green Technology and Innovation, 1(2), 025330013. https://doi.org/ 10.36922/GTI025330013

Riffat, J., Ahadpour Doudran, H., & Samaei, S. R. (2025). AI-enhanced damage detection in jack-up rig legs using an improved modal strain energy index: A numerical, experimental, and digital twin-based approach. Green Technology and Innovation. https://doi.org/10.36922/GTI025340014

Sajedi, S. O., & Liang, X. (2021). Uncertainty-assisted deep vision structural health monitoring. *Computer-Aided Civil and Infrastructure Engineering, 36*(2), 126–142. https://doi.org/10.1111/ mice.12580

Samaei, S. R., & Riffat, J. (2025). Intelligent structural health monitoring of jack-up platform legs using high-density sensor networks, real-time digital twins, and machine learning-based damage detection. *Future Cities and Environment, 11*(1). https://doi.org/10.70917/fce-2025-031

Seyed Reza Samaei, Madjid Ghodsi Hassanabad, Mohammad Asadian Ghahfarokhi, Mohammad Javad Ketabdari, (2021). "Numerical and experimental investigation of damage in environmentally-sensitive civil structures using modal strain energy (case study: LPG wharf)". *Int. J. Environ. Sci. Technol.* 18, 1939–1952. https://doi.org/10.1007/s13762-021-03321-2

Seyed Reza Samaei, Madjid Ghodsi Hassanabad, Mohammad Asadian Ghahfarrokhi, Mohammad Javad Ketabdari, (2021). "Detecting damage location and severity in the leg of SA20 jack up rig (Hull 110) using an iterative modal strain energy method," *Journal of xi'an university of architecture & technology, Page No: 213-236,* https://doi.org/10.37896/JXAT2.1/30319

Shi, H., Worden, K., & Cross, E. J. (2018). A regime-switching cointegration approach for removing environmental and operational variations in structural health monitoring. *Mechanical Systems and Signal Processing, 103*, 381–402. https://doi.org/10.1016/j.ymssp.2017.10.040

Tao, F., Zhang, H., Liu, A., & Nee, A. Y. C. (2018). Digital twin in industry: State-of-the-art. *IEEE Access, 6*, 3585–3600. https://doi.org/10.1109/ACCESS.2018.2791563

Wang, L.-X., Liu, H.-B., Chen, Z.-H., Zhang, F., & Guo, L.-L. (2023). Combined digital twin and hierarchical deep learning approach for intelligent damage identification in cable-dome structure. *Engineering Structures, 274*, 115172. https://doi.org/10.1016/j.engstruct.2022.115172

Wang, Q., Zhu, R., Zheng, J., Su, J., & Wang, N. (2022). A method for quantitative damage identification in a high-piled wharf based on modal strain energy residual variability. *Ocean Engineering, 254*, 111314. https://doi.org/10.1016/j.oceaneng.2022.111314

Wang, X.-Y., & Xia, Y. (2022). Knowledge transfer for structural damage detection through re-weighted adversarial domain adaptation. *Mechanical Systems and Signal Processing, 172*, 108991. https://doi.org/10.1016/ j.ymssp.2022.108991

Worden, K., & Manson, G. (2007). The application of machine learning to structural health monitoring. *Philosophical Transactions of the Royal Society A, 365*(1851), 515–537. https://doi.org/10.1098/ rsta.2006.1938

Xu, X., Xiao, F., Zheng, Y., Pan, T., & Hou, C. (2024). Dynamic response characteristics and pile damage identification of high-piled wharves under dynamic loading. *Applied Sciences, 14*(20), 9250. https://doi.org/10.3390/app14209250

Zhang, R. Y., Meng, L.-B., Mao, Z., & Sun, H. (2021). Spatiotemporal deep learning for bridge response forecasting. *Journal of Structural Engineering, 147*(9), 04021108. https://doi.org/10.1061/ (ASCE)ST.1943-541X.0003022

Zheng, Y., Xiao, F., Zhang, R., Pan, T., Lan, X., Xu, X., & Hou, C. (2024). Experimental investigation and damage identification of high-pile wharf framed bents under horizontal impact loads. *Sensors, 24*(2), 563. https://doi.org/10.3390/s24020563

Zhu, R., Wang, Q., Zheng, J., Su, J., & Wang, N. (2023). Damage detection of foundation pile in high-pile wharf based on statistical high-order moment of dynamic response under wave excitation. *Ocean Engineering, 283*, 115180. https://doi.org/10.1016/j.oceaneng.2023.115180