Journal of *Marine Science and Engineering*

MDPI

Article

Unsupervised Screening and Lightweight Cross-Modal Classification for Damage Identification in Offshore Jacket Structures

Wenxiang Chen, Xinkai Zhao, Jinqi Wang and Zheliang Fan *

School of Chemical Engineering, Ocean and Life Sciences, Dalian University of Technology, Panjin 124221, China; cwx314419@mail.dlut.edu.cn (W.C.); zhaoxinkai@mail.dlut.edu.cn (X.Z.); wangjinqi@mail.dlut.edu.cn (J.W.)
* Correspondence: fanzl@dlut.edu.cn

Abstract

Damage identification in offshore jacket structures requires state screening when damaged samples are scarce and local damage classification under measurement noise. We propose a framework that combines unsupervised screening with a lightweight bimodal classifier. An autoencoder learns undamaged response features from undamaged data alone and provides initial damage screening. The bimodal classifier combines acceleration time histories and mode-shape curvature difference (MSCD) images through bidirectional cross-modal attention and gated fusion. Beam simulations achieved 92.26 ± 1.41% accuracy at 15% Gaussian noise, exceeding direct concatenation by 3.63 percentage points. In scaled jacket tests, the autoencoder detected 98.58% of damaged samples. With 5% low-frequency colored perturbation, the bimodal classifier achieved 100% accuracy across seven structural states on a bootstrap-resampled test set. The classifier contains 3.44 million parameters and required approximately 12 ms per sample on the tested CPU.

Keywords: offshore jacket structure; structural damage identification; autoencoder; mode-shape curvature difference; cross-modal attention; lightweight neural network

1. Introduction

Offshore jacket structures are widely used in offshore oil and gas development and as wind-turbine supports. Cyclic environmental loading and seawater corrosion can cause local damage in their structural members [1,2]. Underwater components are difficult to inspect, and damage may develop between periodic surveys. Vibration-based structural health monitoring (SHM) assesses structural condition through changes in dynamic responses [3,4]. Full-scale tests at West Sole WE compared ambient vibration responses before and after progressive member cutting ahead of platform removal [5]. Local underwater measurements detected partial severance, while global above-water measurements detected complete severance. An actual brace-severance event on a North Sea jacket was also detected from changes in structural response under ambient wave loading [6]. Reliable identification requires distinguishing damage-induced changes from measurement noise, which may also be amplified by features sensitive to local stiffness degradation.

Research linking vibration responses to damage states generally follows model-driven or data-driven approaches. Model-driven methods infer stiffness changes through finite-element model updating, but their results are sensitive to modeling errors and measurement noise [7,8]. Data-driven methods learn damage features from response data [9]. They have progressed from signal indices combined with neural networks to convolutional,

Check for updates

Academic Editors: Bin Liu and Atilla Incecik

Received: 30 September 2026
Revised: 19 October 2026
Accepted: 3 November 2026
Published: 6 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

recurrent, and attention-based architectures. Studies of jacket structures have demonstrated damage localization, damage quantification, and condition assessment under variable excitation [2,10–12]. However, supervised models require labeled damaged samples, which are scarce in operational structures [13,14]. Simulated samples also cannot fully represent differences in measured data [15]. Unsupervised methods can establish a baseline from undamaged data but usually provide only anomaly decisions. They therefore struggle to support label-free initial screening and later classification of specific states [16–19].

Vibration responses of offshore structures contain both temporal evolution and spatial distribution information, but most existing methods emphasize only one representation. Convolutional, recurrent, and attention-based models can extract dynamic changes from acceleration time histories. However, spatial differences caused by local damage usually require additional features. Image-based methods convert transmissibility, time–frequency features, or other monitoring indices into images and use convolutional networks to identify damage patterns [20,21]. These representations typically do not preserve the original temporal order explicitly. A single time-history or image input therefore cannot capture both dynamic processes and local spatial changes. As measurement noise increases, this information loss further impairs the identification of subtle damage.

Multimodal learning provides a route for jointly modeling different types of monitoring information [22]. Previous studies combined multilayer perceptron (MLP)-derived temporal features with ResNet-50 image features for damage identification in high-pile wharves. Other work paired time–frequency maps with raw signals for sensor anomaly diagnosis. Recurrence plots have also been combined with recurrent attention networks for structural damage identification [21,23,24]. These studies demonstrate the value of jointly modeling heterogeneous information. However, direct concatenation relies mainly on the classification head to learn relationships between branches. It cannot selectively exchange complementary information before fusion. High-capacity image backbones and complex fusion modules also increase parameter counts and computational cost. Existing studies emphasize identification accuracy but rarely evaluate branch interaction under noise together with model efficiency.

The transition from offline damage identification to long-term monitoring is also constrained by edge-deployment conditions. Edge computing can analyze data near the source, reducing raw-data transmission and response time. Recent studies have implemented edge-based structural condition assessment in wireless smart-sensor networks [25]. However, this research remains limited. Constraints on processing speed, storage, and energy consumption make complex deep models difficult to deploy directly on field devices. Many damage-identification studies report classification accuracy but do not jointly assess parameter count, computation, and measured hardware latency. Models intended for edge deployment must therefore preserve identification performance, control computational scale, and quantify operating efficiency using measured latency.

We propose a jacket damage-identification framework that combines unsupervised screening with lightweight bimodal classification. The framework addresses scarce damage labels, incomplete use of noisy temporal-spatial information, and high deployment cost. The MSCD feature is motivated by the relationship between mode-shape curvature and flexural rigidity. We also analyze noise propagation through the second difference while retaining acceleration time histories without spatial differencing. During initial monitoring, an autoencoder (AE) is trained only on undamaged data for anomaly screening. After damaged samples are confirmed and accumulated, the bimodal classifier identifies specific states. Building on established network components, the classification contribution concerns task-specific cross-modal interaction and adaptive mixing of original and attention-refined features. Simply supported beam simulations include separate attention

and gating ablations to distinguish their contributions to accuracy and stability under noise. Parameter count, operations, and CPU inference time quantify computational cost. A scaled welded-jacket experiment evaluates screening and classification under connection constraints, including a controlled low-frequency perturbation test for the classifier.

2. Damage Identification Method

2.1. Damage Features and Noise Sensitivity

Structural damage is defined here as a physical change that reduces the stiffness or load-carrying capacity of a jacket member. Typical member damage mechanisms include fatigue cracking under cyclic loading, corrosion-induced wall thinning or pitting, and local deformation or fracture under accidental or extreme loading. Corrosion-related section loss can increase local stresses and accelerate fatigue-crack growth [26]. Such damage can redistribute loads and change local strain magnitudes and principal directions, as well as member rigidity.

These changes may develop gradually through corrosion and stable crack growth, or occur abruptly when a member fractures [6]. Here, numerical stiffness reductions and machined slots create prescribed damage states for identification from acceleration responses and modal features.

The mode-shape difference (MSD) describes amplitude changes, whereas the mode-shape curvature difference (MSCD) emphasizes local variation between adjacent sampling locations. Same-order mode shapes are sign-aligned and normalized before differencing. MSD is defined as:

$$\Delta\phi_i = \phi_i^d - \phi_i^h,\ \Delta\phi \in \mathbb{R}^{N_s} \tag{1}$$

Here, the superscripts h and d denote the undamaged and damaged states, respectively, and N_s is the number of sampling locations. Direct subtraction avoids spatial differentiation, but small damage-induced amplitude changes may be masked by noise. Under Euler–Bernoulli beam theory, curvature is related to flexural rigidity through:

$$\kappa(x) = \phi''(x) \propto \frac{M(x)}{EI(x)} \tag{2}$$

Here, $M(x)$ is the bending moment and $EI(x)$ is the flexural rigidity. Local stiffness loss may therefore produce a curvature anomaly. For equally spaced measurements, curvature at an interior point is estimated by the central difference:

$$\kappa_i = \phi_i'' \approx \frac{\phi_{i+1} - 2\phi_i + \phi_{i-1}}{\Delta x^2} \tag{3}$$

Here, Δx is the distance between adjacent locations. MSCD compares the curvature estimates of the same modal order:

$$\Delta\kappa = \kappa^d - \kappa^h, \Delta\kappa \in \mathbb{R}^{N_s} \tag{4}$$

In the jacket experiment, ten locations are equally spaced along each analyzed member and numbered consecutively in spatial order. Differences are calculated within each member, using its corresponding undamaged reference. With a local axial coordinate s and spacing h_b for member b, $s_{b,i} = s_{b,1} + (i-1)h_b$. The implementation omits division by the squared spacing. Because the locations remain unchanged between structural states, the implemented MSCD and the central-difference estimate in physical coordinates satisfy:

$$\Delta\tilde{\kappa}_{b,i} = {h_b}^2 \Delta\hat{\kappa}_{b,i} \tag{5}$$

Here, $\Delta\tilde{\kappa}_{b,i}$ is the implemented feature and $\Delta\hat{\kappa}_{b,i}$ is the MSCD estimate obtained using the physical spacing. Their scale factor is constant within each member, but may differ

between members. The curvature interpretation applies to the sampled mode-shape component; its relation to flexural rigidity requires a transverse bending component.

Second differencing combines measurement errors with weights of 1, −2, and 1. Assume zero-mean errors that are independent across sampling locations and structural states. With constant variances $(\sigma_\varepsilon^d)^2$ and $(\sigma_\varepsilon^h)^2$ within each state, the noise variance of the MSCD estimate in physical coordinates is:

$$\mathrm{Var}\left(\eta_{(\mathrm{MSCD,i})}\right) = \frac{6\left(\left(\sigma_\varepsilon^{\mathrm{d}}\right)^2 + \left(\sigma_\varepsilon^{\mathrm{h}}\right)^2\right)}{\mathrm{h}^4} \tag{6}$$

Here, h is the physical sampling spacing. For equal error variances in the two states, this variance is twice that of a single-state curvature estimate. Omitting the squared-spacing denominator multiplies the variance by h^4, giving $6((\sigma_\varepsilon^d)^2 + (\sigma_\varepsilon^h)^2)$ for the implemented feature. This rescaling changes both signal and noise amplitudes and does not itself improve the signal-to-noise ratio.

Reversing a member's numbering reverses the difference sequence without changing values at corresponding physical locations. A fixed spatial order is used for image generation. Adjacent interior values are copied to the endpoints to maintain input length; these are padding values rather than independent endpoint-curvature estimates. The simulations and experiments use opposite subtraction signs, with a consistent convention within each separately trained domain.

2.2. Overall Framework

As shown in Figure 1, damage identification comprises anomaly screening and state classification. The screening stage combines MSD and MSCD to describe the undamaged baseline through global amplitude changes and local differences. The classification stage combines acceleration time histories with MSCD images, using temporal patterns to complement the local spatial representation. Screening can begin with only undamaged data, whereas classification identifies specific states after confirmed damage labels have accumulated.

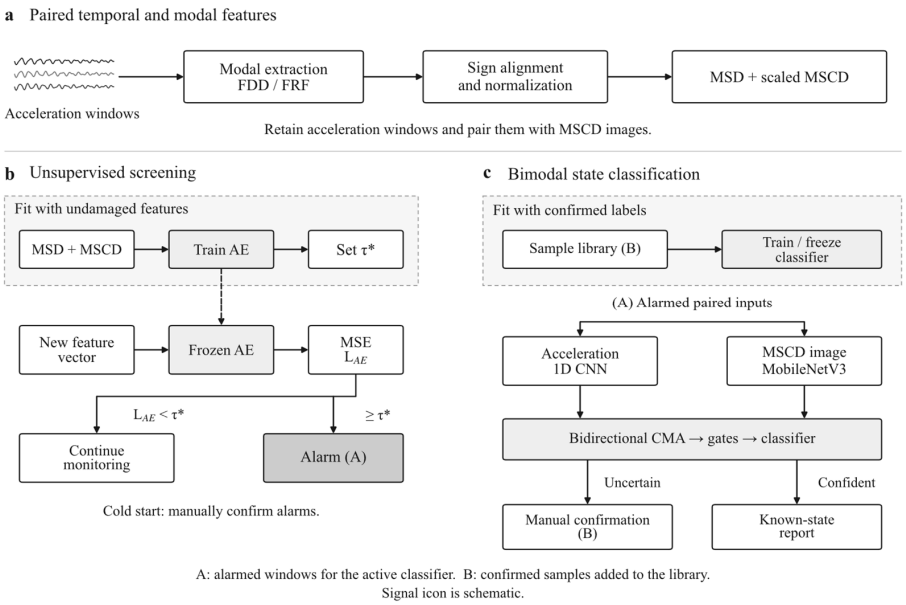


Figure 1. Overall framework of the proposed system.

1. Undamaged-data screening. The autoencoder learns the baseline MSD and MSCD distributions and issues an alarm when the reconstruction error exceeds a threshold.

2. Confirmation and sample accumulation. Confirmed initial alarms become labeled samples. After the classifier is activated, high-confidence predictions for known damage generate reports, while uncertain samples are referred for manual review. Newly confirmed classes are added to the sample library for subsequent training.
3. Bimodal classification. The classifier is trained on confirmed samples to identify damage locations and severities represented by the labels.

2.3. Unsupervised Screening with the Autoencoder

As shown in Figure 2, the AE learns joint MSD and MSCD variations in the undamaged state through a low-dimensional bottleneck. Training keeps reconstruction errors low for undamaged samples. Inputs that deviate from this baseline are screened using their reconstruction errors. This decision uses the undamaged response as its reference and requires no prior collection of damage samples. The encoder and decoder are represented as:

$$\mathbf{h} = \mathrm{f}_{\mathrm{enc}}(\mathbf{x}), \quad \hat{\mathbf{x}} = \mathrm{f}_{\mathrm{dec}}(\mathbf{h}) \tag{7}$$

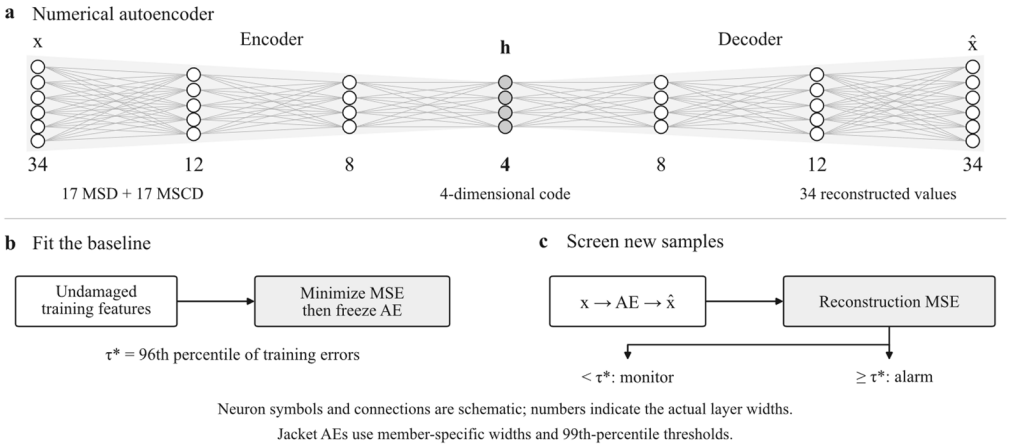


Figure 2. Autoencoder architecture used in the numerical study.

Here, $\mathbf{x}$ is the input feature vector. In the numerical study, $\mathbf{x}$ contains the MSD and MSCD sequences calculated from the preprocessed acceleration responses. $\mathbf{h} \in \mathbb{R}^d$ is the d-dimensional latent code, and $\hat{\mathbf{x}}$ is the reconstructed output. The encoder maps the input to a low-dimensional latent code through fully connected layers. Let l denote the layer index and $n^{(l)}$ the width of layer l. The hidden representation at encoder layer l is:

$$\mathbf{h}^{(\mathrm{l})} = \sigma\left(\mathbf{W}^{(\mathrm{l})}\mathbf{h}^{(\mathrm{l}-1)} + \mathbf{b}^{(\mathrm{l})}\right), \mathrm{l} = 1, \ldots, \mathrm{L}, \mathbf{h}^{(0)} = \mathbf{x} \tag{8}$$

Here, $\mathbf{W}^{(l)} \in \mathbb{R}^{n^{(l)} \times n^{(l-1)}}$ and $\mathbf{b}^{(l)}$ are the weight matrix and bias of layer l, respectively, and σ is the nonlinear activation function. For the numerical AE, the layer widths satisfy $n^{(0)} > n^{(1)} > \ldots > n^{(L)} = d$, compressing the high-dimensional input into a low-dimensional bottleneck. The decoder uses a mirrored architecture that progressively restores the input dimension. The architecture in Figure 2 compresses the 34-dimensional input to 12, 8, and 4 dimensions before applying the mirrored decoder. This design limits model capacity when only baseline samples are available.

The AE minimizes the difference between its input and reconstruction through backpropagation using the mean squared error (MSE) loss:

$$\mathcal{L}_{\mathrm{AE}} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \|\mathbf{x}_{\mathrm{i}} - \hat{\mathbf{x}}_{\mathrm{i}}\|_2^2 \tag{9}$$

Here, $\mathbf{x}_i$ and $\hat{\mathbf{x}}_i$ are input i and its reconstruction, respectively. The numerical AE uses the 96th percentile of reconstruction errors from undamaged training samples as

the threshold τ^*. Experimental thresholds are set separately for each member. During screening, test features are passed through the trained AE to obtain the reconstruction error $\mathcal{L}_{AE}$. If $\mathcal{L}_{AE} \geq \tau^*$, an initial alarm is triggered and the bimodal model compares the sample with the library of known classes. Manual review begins when classification confidence does not reach the specified threshold.

2.4. Lightweight Bimodal Model

As shown in Figure 3, the bimodal model comprises temporal encoding, spatial encoding, and feature fusion. Acceleration time histories preserve response amplitude, frequency, and temporal variation. MSCD images represent scaled curvature differences along the spatially ordered sampling locations of the analyzed beam or jacket member. A 1D CNN and MobileNetV3 compress the two inputs into low-dimensional embeddings. Bidirectional attention links the branches, and gates adjust the mixture of each original representation and its interaction-derived representation.

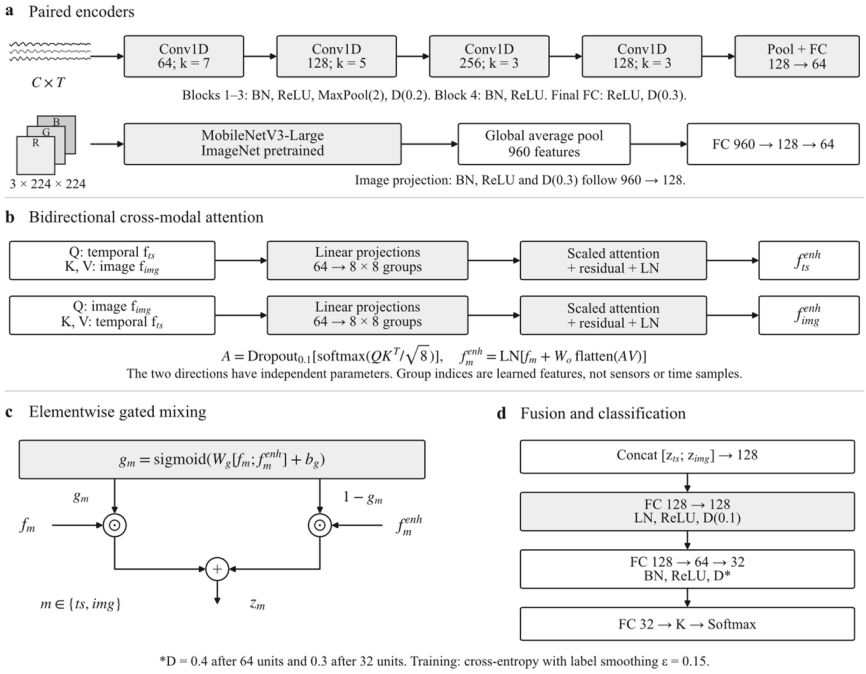


Figure 3. Architecture of the bimodal model. CMA, cross-modal attention.

2.4.1. One-Dimensional Convolutional Neural Network

The temporal branch organizes each acceleration window with measurement points as channels and samples along the time dimension. A one-dimensional convolutional neural network (1D CNN) then extracts a response representation [27]. Let one window be $X \in \mathbb{R}^{C \times T}$, where C is the number of input channels and T is the number of samples. One-dimensional kernels slide along the time axis to extract local waveforms and combine information across input channels. Shared weights allow the same response pattern to be recognized at different temporal positions within a window.

$$u_j^{(l)}(t) = \sum_{c=1}^{C_{l-1}} \sum_{\tau=0}^{k_l-1} w_{jc\tau}^{(l)} h_c^{(l-1)}\left(t + \tau - p_l\right) + b_j^{(l)} \tag{10}$$

$$a_j^{(l)}(t) = \mathrm{ReLU}\left(\mathrm{BN}\left(u_j^{(l)}(t)\right)\right) \tag{11}$$

The network contains four one-dimensional convolutional layers with a stride of 1. Their kernel sizes are 7, 5, 3, and 3, and their output-channel counts are 64, 128, 256, and

128, respectively. After activation output $a^{(l)}$, each of the first three convolutional blocks applies twofold max pooling and dropout of 0.2. This yields $h^{(l)}$ for the next block, while the fourth block uses $h^{(4)} = a^{(4)}$. Successive convolution and pooling expand the temporal range represented by later features while shortening the sequence.

To summarize responses of different lengths as fixed-dimensional representations, adaptive average pooling is applied along time after the fourth block, followed by a 64-dimensional projection:

$$\bar{h}_j = \frac{1}{T'} \sum_{t=1}^{T'} h_j^{(4)}(t) \tag{12}$$

$$f_{ts} = \text{Dropout}_{0.3}\left[\text{ReLU}\left(W_p \bar{h} + b_p\right)\right] \tag{13}$$

Here, T' is the temporal length of the fourth-block output. $\bar{h} \in \mathbb{R}^{128}$ is the pooled feature vector, while $W_p \in \mathbb{R}^{64\times128}$ and $b_p \in \mathbb{R}^{64}$ are the projection weight and bias, respectively. The resulting f_{ts} interacts with the 64-dimensional spatial embedding from the image branch in the subsequent feature groups.

The temporal branch forms a window-level response representation through successive local filtering. Early convolutions extract short-term waveforms, while later convolutions combine patterns over wider temporal ranges after pooling. Average pooling summarizes the response intensity of each feature channel. Because the input undergoes no spatial second difference, it retains dynamic information not represented in the curvature image and complements the joint classification. The beam simulations use 5–34 Hz band-pass filtering for the acceleration branches. The scaled-jacket classifier uses the impact-response windows described in Section 4.1.

2.4.2. MobileNetV3

Each MSCD image uses the measurement-point index on the horizontal axis and differential amplitude on the vertical axis. Damage-related changes therefore appear as a polyline. Local peaks and valleys describe abrupt changes between neighboring points, while broader fluctuations reflect concurrent changes at several locations. Two-dimensional convolutions first extract morphological features near the polyline and then combine them into a global curve representation. The horizontal coordinate follows the spatial order along the analyzed beam or jacket member. The image therefore represents its scaled MSCD distribution, rather than a two-dimensional geometric projection of the complete jacket.

The image branch uses ImageNet-pretrained MobileNetV3-Large with 224 × 224 RGB MSCD images as input [28]. Its 960-dimensional pooled output is projected successively to 128 and 64 dimensions, matching the temporal branch. The intermediate projection uses batch normalization, *ReLU*, and dropout of 0.3. This compact embedding supports feature-group interaction and reduces the cost of directly fusing high-dimensional image features. As illustrated in Figure 4, depthwise separable convolutions in the backbone separate spatial filtering from channel mixing and are expressed as:

$$\mathbf{y} = \text{Conv}_{1\times1}(\text{DWConv}_{k\times k}(\mathbf{x})) \tag{14}$$

Here, x and y are the input and output feature maps, respectively, and k is the spatial kernel size.

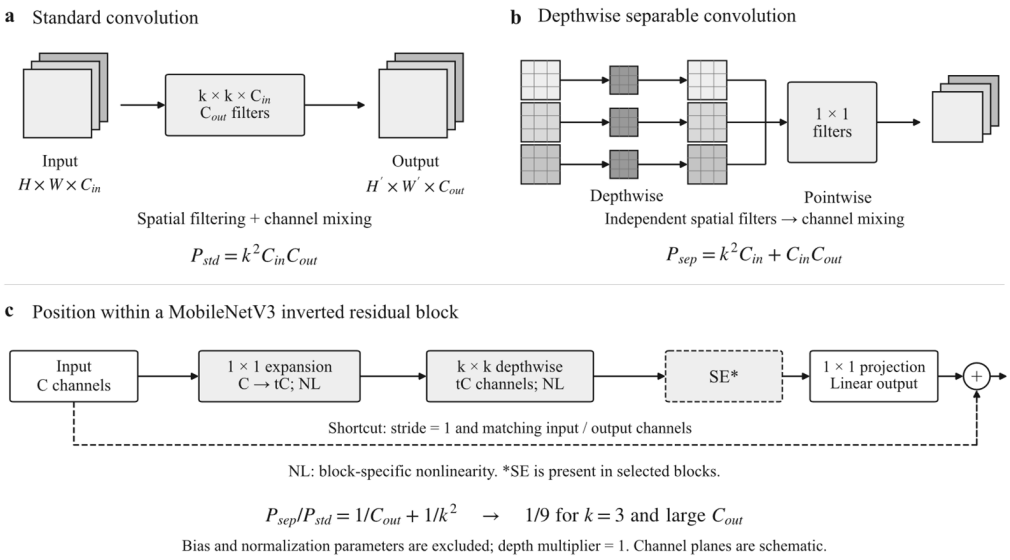


Figure 4. Principle of depthwise separable convolution.

The difference in parameter count motivates the lightweight design. The total parameters in a standard $k \times k$ convolution and a depthwise separable convolution are:

$$\begin{aligned} P_{std} &= C_{in} \cdot C_{out} \cdot k^2 \\ P_{dw} &= C_{in} \cdot k^2 + C_{in} \cdot C_{out} \end{aligned} \tag{15}$$

Here, C_{in} and C_{out} are the numbers of input and output channels, respectively. When the number of output channels is sufficiently large and the kernel size is k = 3, depthwise separable convolution uses approximately $\frac{1}{9}$ of the standard-convolution parameters. This factorization preserves spatial filtering and channel combination while reducing convolutional parameters.

The inverted residual block first expands the channels to enrich the intermediate representation and then compresses the features through a linear projection. Channel attention adjusts the responses of different feature channels according to the input [28,29]. The compact image branch can thus combine local peaks and valleys with the overall curve shape. It then produces a spatial representation with the same dimension as the temporal embedding.

2.4.3. Cross-Modal Attention and Gated Fusion

Direct concatenation presents the two branch representations in parallel to the classification head, which learns their relationship in subsequent mappings. Here, branch interaction occurs before concatenation. Each branch uses features from the other as supplementary information and forms a refined representation through bidirectional attention. Gates then control the proportions of the original and refined features. The classification head therefore receives a joint representation after cross-branch information exchange.

Each branch outputs a 64-dimensional embedding. Using scaled dot-product attention [30], the linear query, key, and value projections are reshaped from B × 64 to B × 8 × 8, where B is the batch size. Attention exchanges information between eight learned feature groups from the two branches:

$$\mathrm{Attn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \mathrm{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\mathbf{V} \tag{16}$$

Here, Q, K, and V are the query, key, and value matrices, respectively. d_k equals 8, and $\sqrt{d_k}$ is the scaling factor. Each sample produces an 8 × 8 attention matrix normalized along

the key-group dimension, with attention dropout of 0.1. The weighted values are flattened to 64 dimensions, projected, added to the original query embedding, and layer-normalized. Both directions apply this operation across learned feature groups. The group indices do not correspond to temporal samples or physical measurement locations.

Adaptive gate weights then mix the attention-refined features:

$$\begin{aligned} \mathbf{f}_{\mathbf{ts}}^{\mathbf{out}} &= \mathbf{f}_{\mathbf{ts}} \odot \mathbf{g}_{\mathbf{ts}} + \mathbf{f}_{\mathbf{ts}}^{\mathbf{enh}} \odot \left(\mathbf{1} - \mathbf{g}_{\mathbf{ts}}\right) \\ \mathbf{f}_{\mathbf{img}}^{\mathbf{out}} &= \mathbf{f}_{\mathbf{img}} \odot \mathbf{g}_{\mathbf{img}} + \mathbf{f}_{\mathbf{img}}^{\mathbf{enh}} \odot \left(\mathbf{1} - \mathbf{g}_{\mathbf{img}}\right) \end{aligned} \tag{17}$$

Here, f_{ts} and f_{img} are the original branch features. f_{ts}^{enh} and f_{img}^{enh} are the attention-refined features, while g_{ts} and g_{img} are the temporal and spatial gate weights. Each gate concatenates the original and refined features within its branch. A 128-to-64 linear layer and a sigmoid function then generate the weights. ⊙ denotes elementwise multiplication.

Each gate operates on feature elements within its branch, allowing sample-specific combinations of original and refined representations. This allows cross-modal interaction to be moderated while retaining original branch information. The branch outputs are concatenated and passed through a 128-to-128 projection, layer normalization, ReLU, and dropout of 0.1. The classifier uses hidden layers of 64 and 32 dimensions. They apply batch normalization, ReLU, and dropout rates of 0.4 and 0.3 before the output layer produces class probabilities.

The classifier is trained with label-smoothed cross-entropy loss:

$$\mathcal{L}_{\mathbf{cls}} = -\sum_{\mathbf{c}=\mathbf{1}}^{\mathbf{K}} \tilde{\mathbf{y}}_{\mathbf{c}} \mathbf{log} \mathbf{p}_{\mathbf{c}}, \ \tilde{\mathbf{y}}_{\mathbf{c}} = (1 - \boldsymbol{\varepsilon}) \mathbf{y}_{\mathbf{c}} + \frac{\boldsymbol{\varepsilon}}{\mathbf{K}} \tag{18}$$

Here, K is the number of classes, y_c is the one-hot label, and $\tilde{y}_c$ is the smoothed label. The smoothing coefficient is ε = 0.15. Label smoothing regularizes supervised training [31], whereas the AE is trained separately using only undamaged data.

3. Numerical Simulation and Model Evaluation

3.1. Simply Supported Beam Model and Damage Cases

The curvature relation and noise analysis in Section 2.1 motivate numerical evaluation of MSCD after modal extraction and classification. We used the simply supported beam model in Figure 5 to compare controlled damage cases across Gaussian noise levels.

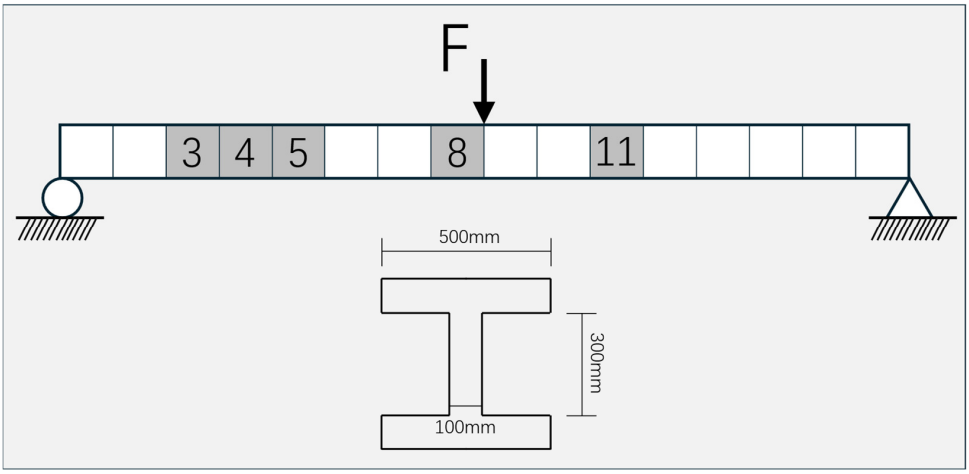


Figure 5. Numerical model of the simply supported beam.

The beam span is 8 m and is discretized into 16 BEAM188 elements. The elastic modulus, Poisson's ratio, and density are 200 GPa, 0.3, and 7850 kg/m^3, respectively. Rayleigh

damping is used with a damping ratio of 0.02. A total of 17 acceleration measurement points are uniformly distributed along the beam. Random excitation with an amplitude range of 0–4000 N is applied, and vertical acceleration is sampled at 500 Hz. Local damage is simulated by reducing the elastic modulus of selected elements. Table 1 contains five single-damage and four combined-damage cases. They examine the effects of location, severity, and overlapping responses from multiple damaged regions.

Table 1. Damage cases in the numerical study.

ID	Damaged Element(s)	Damage Severity (%)
D1	3	30
D2	5	20
D3	5	30
D4	5	40
D5	8	30
D6	3, 5	30
D7	5, 8	30
D8	3, 8	30
D9	4, 11	30

Frequency-domain decomposition (FDD) is used to estimate the mode shape from each acceleration window [32]. Only the first mode is retained, and its amplitudes are ordered by physical measurement points 1–17.

Each MSCD is plotted as a 224 × 224 RGB image with a 17-point blue polyline on a white background. No spatial interpolation or axes are used, and all images share the same vertical-axis range. Each acceleration time history produces 296 windows using 500 samples per window and a step of 100 samples. The windows are split chronologically at a 6:2:2 ratio. Each class contains 177 training, 59 validation, and 60 test windows. For image generation, the ith successfully extracted damaged mode shape is compared with the corresponding ith undamaged mode shape. Images are matched to acceleration windows by filename index. For AE vector features, the reference is the sign-aligned mean mode shape of the quality-retained undamaged training windows. This organization retains inter-window dependence and the shared reference information.

Gaussian noise is scaled by the standard deviation of the complete time history from each sensor. Noise levels of 5–15% denote the ratio of noise to signal standard deviation. For the model comparison, each configuration is trained five times, and accuracy is reported as the mean ± standard deviation. The main classifier uses a batch size of 64 and up to 100 epochs. Early stopping uses a validation-loss patience of 30 epochs. AdamW learning rates are 2×10^{-4} for non-image base parameters, 1.5×10^{-4} for the classification head, and 1×10^{-5} for the image branch. Weight decay is 2×10^{-3}, and cosine scheduling reduces the learning rate to 10^{-6}.

3.2. Damage Localization and Screening

3.2.1. Damage Localization Using Curvature Differences

Figure 6 shows MSCD images for selected damage cases. Distinct local extrema occur near the prescribed damaged elements, whereas changes in undamaged regions remain relatively small. Local stiffness loss changes the relationship between adjacent mode-shape amplitudes. The second difference emphasizes this change as peaks and valleys, improving the distinguishability of damage locations. For adjacent combined damage, responses from different damaged regions overlap and local peaks may merge, reducing spatial resolution. A single extremum cannot then describe the complete damage configuration.

Peak and valley locations, relative amplitudes, and the overall curve shape jointly distinguish the states.

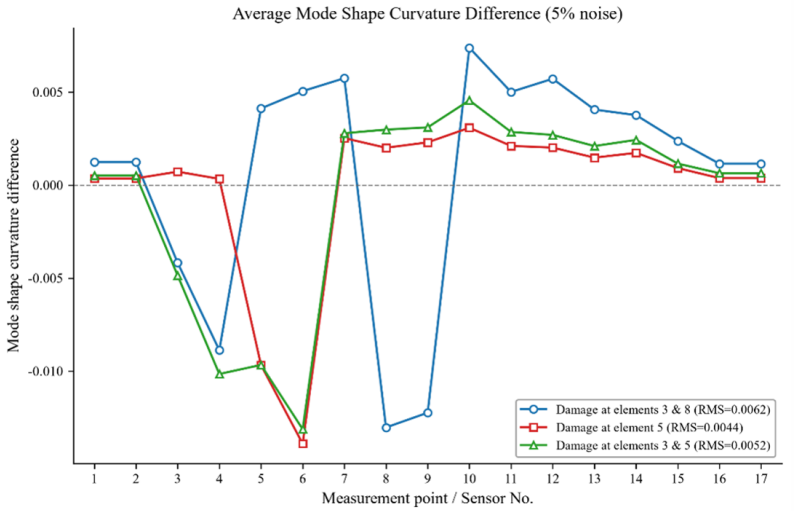


Figure 6. MSCD distributions for representative damage cases.

3.2.2. Unsupervised Damage Screening

The AE was fitted separately at each noise level using only midpoint undamaged data. Acceleration windows were filtered at 5–34 Hz before FDD, with 333-point Welch segments and 166-point overlap. Using these modal estimates, the 34-dimensional input combined MSD and MSCD. The excitation-quality threshold was the 10th percentile of acceleration RMS values from the first 207 undamaged windows. Of the 256 retained undamaged windows, the first 179 formed the training set and the remaining 77 formed the test set. The sign-aligned mean reference mode shape, standardization statistics, and 96th-percentile reconstruction-error threshold were calculated from the training set. All damaged windows passing the same quality threshold were evaluated. Training used Adam with a learning rate of 5×10^{-4}, weight decay of 10^{-3}, and a batch size of 32. Training ran for at most 300 epochs, with an early-stopping patience of 40 epochs. The AE used initialization seed 19 at each noise level. The percentile was selected using midpoint holdout data and then fixed for the shifted-excitation evaluation.

As shown in Figure 7 and Table 2, AE accuracy was 99.92% without noise and 92.48% at 5% noise. Damage recall was 100.00% and 92.36%, respectively. At 10% and 15% noise, recall decreased to 77.61% and 58.02%, while precision remained above 99.8%. Across 0%, 5%, 10%, and 15% noise, false-alarm rates were 2.60%, 3.90%, 3.90%, and 2.60%, respectively. At 15% noise, 990 of 2358 damaged windows were missed, compared with two false alarms among 77 undamaged windows. Noise perturbs modal estimates and enters MSD and MSCD through differencing, increasing the overlap between undamaged and damaged reconstruction errors. The 96th-percentile threshold therefore did not fully separate the two states at higher noise levels. These results support retaining manual review when automated screening is uncertain. Overlapping windows and holdout-based threshold selection limit interpretation to an exploratory evaluation.

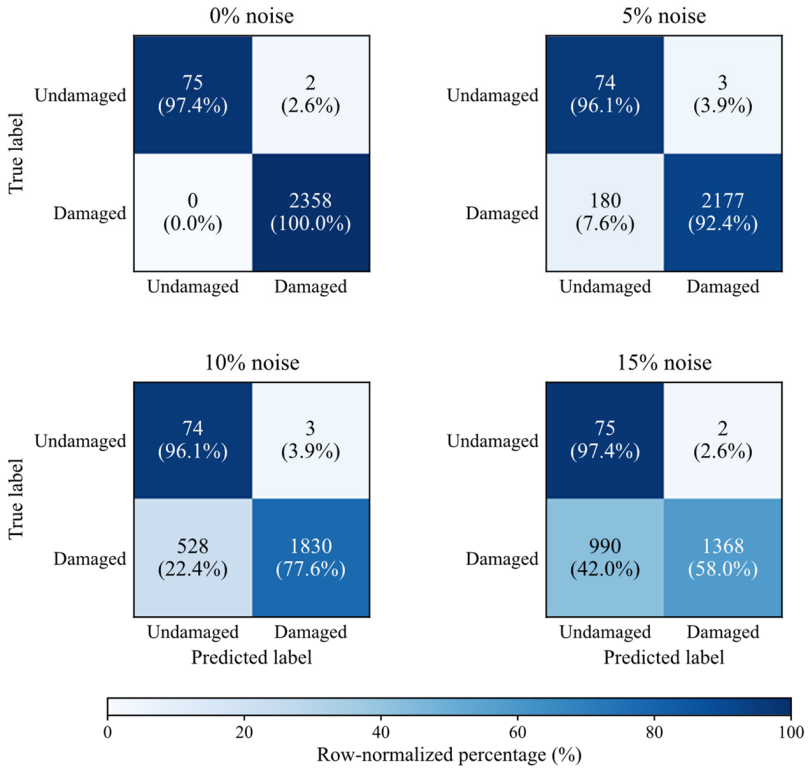


Figure 7. Autoencoder confusion matrices at different noise levels. Cell values give counts and row-normalized percentages.

Table 2. Autoencoder performance at different noise levels.

	Accuracy	Precision	Recall	F1 Score
0%	0.999	0.999	1.000	1.000
5%	0.925	0.999	0.924	0.960
10%	0.782	0.998	0.776	0.873
15%	0.593	0.999	0.580	0.734

Anomaly screening indicates deviation from the undamaged baseline, but the classifier must also handle inputs outside the known labels. To examine this case at 5% Gaussian noise, D1-D8 are treated as eight known damage classes. D9 is treated as unknown and excluded from training. The maximum softmax probability is used as the known-class score [33]:

$$\mathbf{s}(\mathbf{x}) = \max_{\mathbf{k}} \mathbf{p}_{\mathbf{k}}(\mathbf{x}) \tag{19}$$

A sample is accepted as known when $s(x) \geq \tau$ and rejected as uncertain when $s(x) < \tau$. The threshold analysis scans percentiles of the training-set confidence distribution and adopts its 5th percentile: $\tau = 0.7164$.

A threshold of 0.7164 retained 95.00% of the 480 known test samples and rejected 99.66% of the 296 D9 samples, as shown in Figure 8. Known samples generally received high class scores, whereas D9 samples were concentrated at lower confidence. Damage that does not match a known label therefore appears as classification uncertainty. This result supports referring low-confidence samples for review.

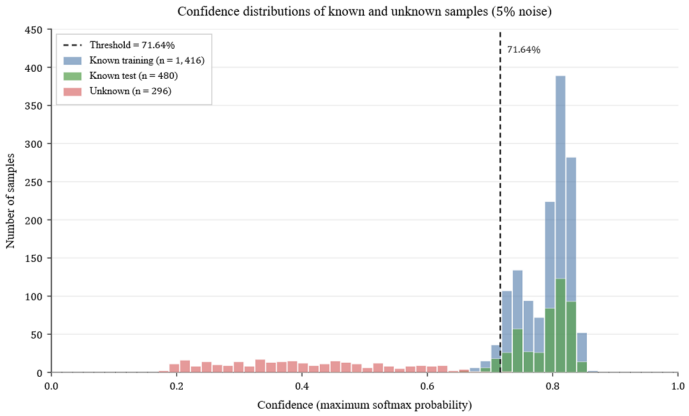


Figure 8. Confidence distributions of known and unknown samples.

3.3. Framework Evaluation at a Different Excitation Location

To examine the effect of excitation location, the load was moved from midspan to the fourth node from the left. The beam properties, excitation time history and amplitude, damping, sampling settings, and damage configuration were unchanged. D9, with a 30% reduction in the elastic modulus of elements 4 and 11, was evaluated under 5% Gaussian noise. The AE and bimodal classifier used midpoint-excitation training data, with initialization seeds of 19 and 314, respectively. Both models retained their original parameters and undamaged references during evaluation.

The response record provided 296 windows of 500 samples with a 100-sample step. The AE used the 333-point Welch setting and 96th-percentile threshold defined in Section 3.2.2. The excitation-quality rule retained 234 windows, corresponding to 79.05% of the record. Figure 9 compares their reconstruction errors with the fixed threshold. The minimum reconstruction error was 2.200, above the threshold of 1.892, and all 234 retained windows triggered an alarm.

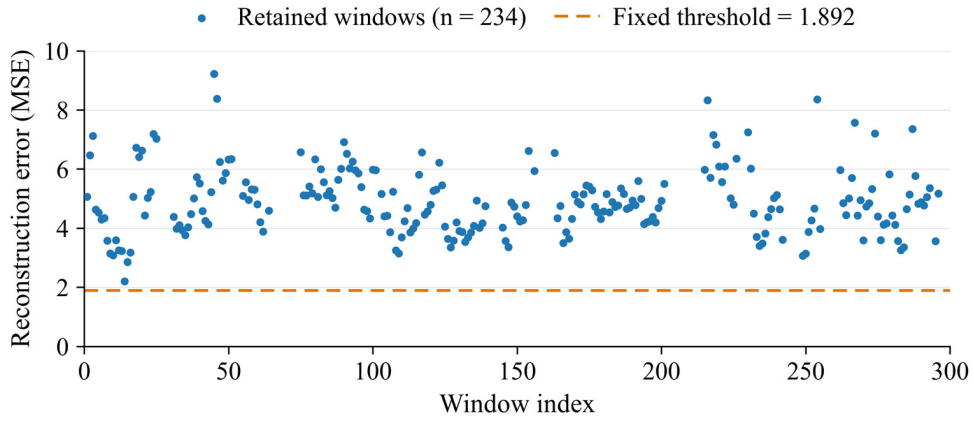


Figure 9. D9 AE reconstruction errors with fourth-node excitation (5% Gaussian noise).

The bimodal classifier used matched acceleration windows and MSCD images, with image scaling and window-specific undamaged references taken from the midpoint data. This nine-class model included D9 during training; the uncertainty test in Section 3.2.2 used a separate eight-class model. Across the complete record, 277 of 296 windows were classified as D9; the remaining 19 were assigned to D4. The classifier therefore achieved 93.58% accuracy under the shifted excitation.

Table 3 summarizes the screening and classification results. For the joint evaluation, each retained window required both an AE alarm and the correct D9 label. This criterion

was satisfied by 226 of 234 windows, giving a conditional success rate of 96.58%. The joint rate refers to the windows retained for AE screening, whereas classifier accuracy uses all 296 windows.

Table 3. D9 screening and classification with shifted excitation (5% Gaussian noise).

Stage	Evaluated Windows	Successful Decisions	Rate (%)
AE screening	234 retained	234 detections	100.00
Bimodal classification	296 total	277 correct	93.58
Joint decision	234 retained	226 joint successes	96.58

The AE detected deviation from the midpoint undamaged baseline, while the bimodal classifier identified the combined damage state in most windows. Together, these results support applying the two-stage framework to the tested excitation location without updating the models or undamaged references.

3.4. Model Comparison and Noise Robustness

Table 4 compares single-branch models (M1–M2), direct concatenation (M3), ShuffleNetV2 [34] and MobileNetV3 configurations (M4–M5), and models without attention (M7) or gating (M8). All models use the same nine-class beam data and chronological split, with their corresponding optimization settings. Acceleration branches share the filtering described in Section 2.4.1.

Table 4. Classification accuracy (%) of candidate and ablated models at different noise levels (mean ± standard deviation across five training runs).

ID	Model	0% Noise	5% Noise	10% Noise	15% Noise
M1	1DCNN	100.00 ± 0.00	99.37 ± 0.43	95.78 ± 1.94	91.04 ± 3.29
M2	MNV3	100.00 ± 0.00	95.26 ± 0.38	73.67 ± 1.44	50.37 ± 0.99
M3	1DCNN-MNV3-Concat	100.00 ± 0.00	99.67 ± 0.30	94.85 ± 1.77	88.63 ± 2.48
M4	1DCNN-SNV2-CrossATT	100.00 ± 0.00	98.30 ± 3.11	94.22 ± 3.38	92.59 ± 2.53
M5	1DCNN-MNV3-CrossATT	100.00 ± 0.00	99.41 ± 0.67	96.11 ± 1.72	92.26 ± 1.41
M7	Without attention	100.00 ± 0.00	99.19 ± 0.75	90.74 ± 5.09	88.59 ± 3.67
M8	Without gating	100.00 ± 0.00	99.63 ± 0.73	97.07 ± 1.72	92.56 ± 3.27

Without added noise, all models achieved 100.00 ± 0.00% accuracy, indicating sufficient classification capacity for the prescribed damage cases. The image branch degraded faster as noise increased. At 15% noise, M2 fell to 50.37% mean accuracy, whereas M1 retained 91.04%. This difference is consistent with the sensitivity of MSCD to modal-estimation errors amplified by second differencing. The filtered temporal input avoids this spatial-difference amplification.

M3 achieved 99.67% mean accuracy at 5% noise but fell to 88.63% at 15% noise, below M1. M5 reached 92.26% at 15% noise, 3.63 percentage points above M3. Removing attention (M7) reduced mean accuracy from 96.11% to 90.74% at 10% noise and from 92.26% to 88.59% at 15% noise. The corresponding standard deviations increased from 1.72 to 5.09 and from 1.41 to 3.67 percentage points. These comparisons support the contribution of cross-modal interaction to accuracy and stability under stronger noise.

At 15% noise, gating reduced the standard deviation from 3.27 (M8) to 1.41 (M5) percentage points, a decrease of 56.9%. This indicates more consistent classification across training runs under strong noise. Gating adaptively combines original and attention-

refined features within each branch, which may limit the influence of perturbed cross-modal interactions.

At 15% noise, M4 and M5 achieved similar mean accuracies (92.59% and 92.26%), whereas M5 had a lower standard deviation (1.41 versus 2.53 percentage points). Considering both classification accuracy and stability, M5 was selected as the main model.

3.5. Computational Efficiency

Complexity is measured by parameters and operations, and latency on an Intel i7-13650HX CPU. M5 contains 3.44 million parameters and requires 12.18 ms per sample. Relative to MLP-ResNet50 benchmark M6 [24], M5 uses approximately 86% fewer parameters and is approximately 2.9 times faster. Relative to direct concatenation M3, it adds only 0.05 million parameters and 0.46 ms of latency. Mean accuracy at 15% noise increases by 3.63 percentage points. As listed in Table 5, ShuffleNetV2 configuration M4 has approximately half the parameters of M5 and fewer operations, but its measured latency was 12.22 ms, compared with 12.18 ms for M5. The smaller backbone therefore did not yield lower measured CPU latency.

Table 5. Computational characteristics of the candidate models.

ID	Model	Parameters (Million)	Operations (G)	Inference Latency (ms)	Samples per Second
M3	1DCNN-MNV3-Concat	3.39	0.4285	11.72	85.40
M4	1DCNN-SNV2-CrossATT	1.73	0.2882	12.22	81.96
M5	1DCNN-MNV3-CrossATT	3.44	0.4286	12.18	82.18
M6	MLP-ResNet50	24.73	8.1767	35.52	28.19

Considering accuracy, stability across training runs and CPU latency, the MobileNetV3-based M5 configuration offered the best overall balance among the evaluated models. Including signal preprocessing, FDD-based modal extraction, MSCD image generation and classification, mean processing latency was 31.16 ± 8.13 ms on the same CPU. These measurements excluded physical acquisition, communication and AE screening; end-to-end latency on representative edge devices remains unverified.

4. Experimental Validation on a Scaled Jacket Model

The beam simulation evaluated damage identification across damage cases and noise levels. The jacket experiment extends the beam simulation to members subject to welded-joint and brace-connection constraints, while retaining equally spaced sampling locations along each analyzed member. We therefore introduced slotted-member damage into a scaled welded-jacket model and retrained the models using measured data. The experiment evaluates anomaly screening and predefined-state classification for different members. It extends the numerical analysis of feature use to physical measurements under welded connection constraints.

4.1. Experimental Model and Protocol

A scaled model was designed from a representative offshore jacket platform using elastic-gravity similitude. The rectangular frame in Figure 10 has a total height of H = 1.7 m and base dimensions of 1.8 m × 1.25 m. Each elevation has a rectangular plan. The model is made of galvanized steel with an elastic modulus of E = 206 GPa and a yield strength of f_y = 235 MPa. The columns and braces are circular steel tubes with an outer diameter

of 30 mm and a wall thickness of 2 mm. All members are welded, and the total mass is approximately 50 kg.



Figure 10. Scaled jacket model and damage locations.

A roving impact-hammer method was used to reduce the effect of sensor mass. A DH187E accelerometer with a sensitivity of 9.83 mg/V was fixed at the central node of the bottom level. For each analyzed short member, long member, or brace, an LC02-3A102 impact hammer excited ten equally spaced locations along the same member. The locations were numbered consecutively along its local axial coordinate. Modal features were constructed separately for each member and compared with its corresponding undamaged reference. At each location, 20 valid records were retained after coherence-based screening of 40 impacts. Acceleration was sampled at 1000 Hz. For classification, each frequency-response function (FRF) window contained 100 pre-impact and 2500 post-impact samples. The temporal input comprised 2500 samples from the same window start. The H1 FRF was estimated from the impact records. Its imaginary component at the selected resonance frequency was used as the relative mode-shape amplitude.

Classification data were divided by original impact number before bootstrap resampling. Impacts 1–10, 11–15, and 16–20 formed the training, validation, and test pools, respectively. Incomplete windows, windows crossing recording boundaries, and impacts failing the force-signal quality rule were excluded. This removed one training impact at the second long-member damage location; no other impact among the first 20 was excluded. For each sample and excitation location, one eligible impact was selected as the temporal anchor. Two additional impacts were drawn without replacement from the same pool, and all three were used to estimate the FRF. Anchor selection was balanced across eligible impacts, independently at each location. In each validation and test pool, all five impacts served as anchors 40 times across 200 samples. The ten response segments were stacked in excitation-location order; they were sequential measurements from one fixed accelerometer. Each class contained 300 training, 200 validation, and 200 test samples. The impact pools did not overlap. Reference frequencies, undamaged mode shapes, amplitude references,

and MSCD plotting limits used training data only. Bootstrap combinations reused original impacts and did not constitute independent physical tests.

Slots were machined into short members, long members, and braces to reduce their effective load-bearing area and stiffness. Two slot lengths were used for the short member, as listed in Table 6. The member types represent different connection and load-transfer conditions. The two damage levels in the same short member provide a severity comparison.

Table 6. Damage cases in the scaled-model experiment.

ID	Damage Location	Slot Length (cm)
D1	Short member	5
D2	Short member	6.5
D3	Long member	4
D4	Brace	4

4.2. Autoencoder Screening Performance

The experimental AE uses a 20-dimensional input comprising first-mode MSD and scaled MSCD at ten equally spaced locations along each analyzed member. The short-member encoder widths are 20-18-12-4, while the long-member and brace encoders use 20-24-16-3. All models use mirrored decoders and hidden-layer dropout of 0.1. Before bootstrap resampling, the 20 impacts were randomly divided into two groups. The groups generated the undamaged training and test sets, respectively.

Adam learning rates were 3×10^{-4}, 5×10^{-4}, and 2×10^{-4} for the short member, long member, and brace, respectively. The corresponding maximum epoch counts were 200, 200, and 500. The batch size was 32, weight decay was 10^{-4}, and early-stopping patience for training loss was 30 epochs. For all three members, the alarm threshold was the 99th percentile of reconstruction errors from undamaged training samples.

Figure 11 shows that all damaged samples from the short member and brace were detected. For the long member, 283 of 300 damaged samples were detected, giving a detection rate of 94.3%. False-alarm rates for undamaged short-member, long-member, and brace samples were 0.3%, 16.0%, and 11.3%, respectively. Across all members, 1183 of 1200 damaged samples were detected, and 83 false alarms occurred among 900 undamaged samples. The overall detection rate, false-alarm rate, and accuracy were 98.58%, 9.22%, and 95.24%, respectively. All 17 missed detections came from the long member.

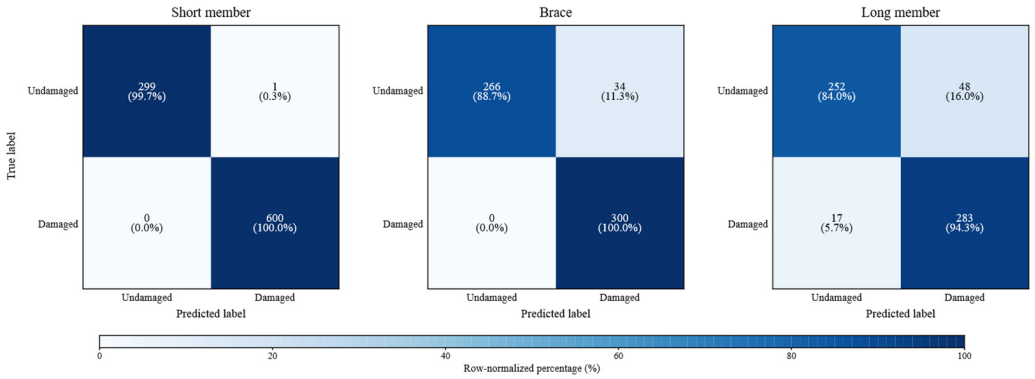


Figure 11. Autoencoder confusion matrices for the three member types.

The long member is welded at both ends and reinforced by a brace connection near midspan. This intermediate restraint changes its bending response relative to a member constrained only at its ends. It may redistribute modal deformation on either side of the

connection and reduce the damage-induced curvature contrast captured at the sampling locations. A weak local change may then be obscured by modal-identification errors amplified through second differencing. The damaged features may remain close to the undamaged distribution, limiting the increase in AE reconstruction error and increasing missed detections. Meanwhile, variability in undamaged modal estimates can broaden the baseline reconstruction-error distribution and produce false alarms at a fixed threshold. These mechanisms may help explain why the long member showed both missed detections and a higher false-alarm rate than the short member and brace.

4.3. Bimodal Damage Classification

The classifier distinguished seven states using ten-channel acceleration inputs with 2500 samples per channel. Training used unperturbed data (seed 314), a batch size of 16, and at most 50 epochs, with early-stopping patience of 10 epochs. AdamW learning rates, weight decay, and label smoothing matched the numerical model. The checkpoint with the lowest validation loss was selected for testing.

Testing used a 5% low-frequency colored perturbation with power spectral density proportional to 1/f between 1 and 20 Hz and zero outside this band. Gaussian random sequences were spectrally shaped and added to the original acceleration records before either classifier input was constructed. Within each continuous recording segment, the zero-mean perturbation RMS was set to 5% of the corresponding undamaged training reference. This reference was the median standard deviation of eligible undamaged training FRF windows at that excitation location. A fixed base seed of 314 generated a reproducible realization for each recording segment. The force records, balanced anchors, FRF impact groups, and training references were unchanged. FRFs, MSCD images, and temporal inputs were recomputed from the perturbed acceleration.

Training loss stabilized near 0.62, while validation loss continued to fluctuate (Figure 12). The checkpoint selected at epoch 30 achieved 100% training accuracy and 99.86% validation accuracy; early stopping occurred at epoch 40.

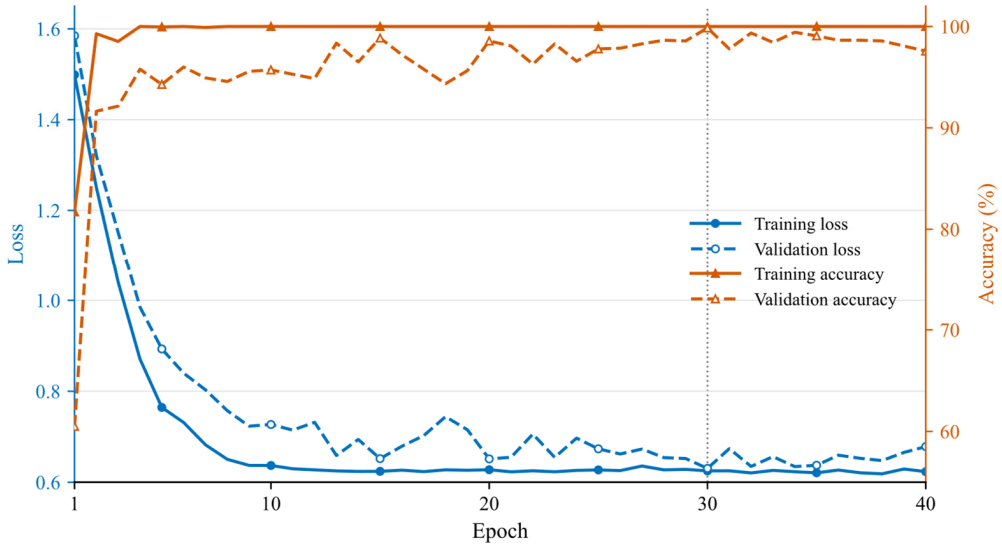


Figure 12. Training and validation loss and accuracy.

The classifier achieved 100% accuracy on 1400 bootstrap-resampled test samples under 5% low-frequency colored perturbation (Figure 13). Macro-averaged F1 was 1.000, and accuracy matched the unperturbed control. The disturbance band lay below the healthy reference FRF peaks.

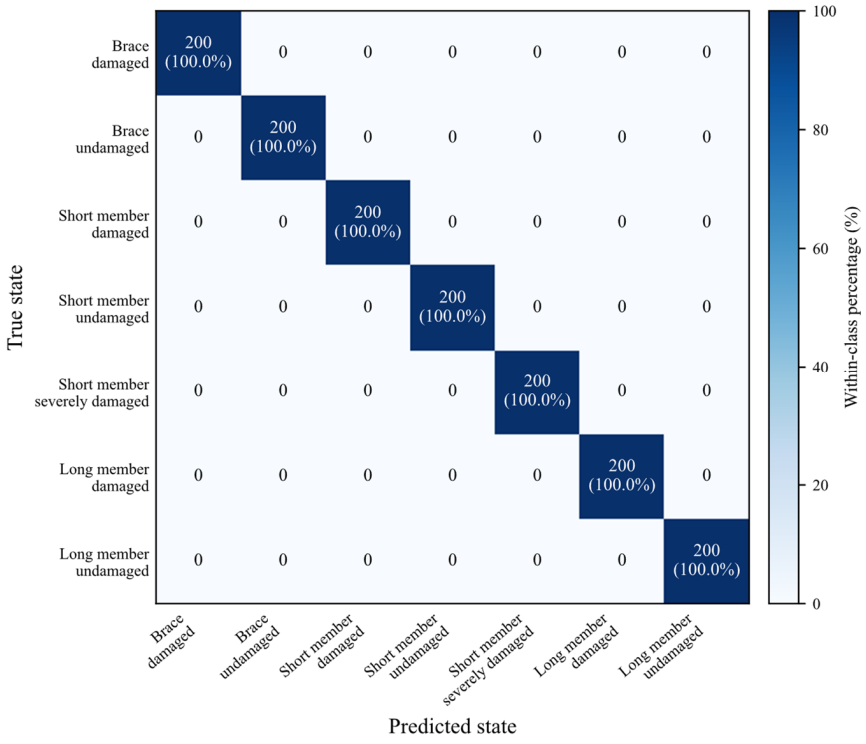


Figure 13. Confusion matrix under 5% low-frequency colored perturbation.

5. Conclusions

We developed a closed-loop framework for scarce damage samples and limited edge resources in structural health monitoring of offshore jacket structures. It integrates unsupervised AE screening, damage confirmation and sample accumulation, and bimodal classification. The framework cold-starts from undamaged data and uses confirmed damage samples to update the classifier. Simply supported beam simulations and scaled-jacket experiments evaluated its performance. The main conclusions are as follows.

- At 15% Gaussian noise, the complete bimodal model achieved 92.26 ± 1.41% accuracy across five training runs, 3.63 percentage points above direct concatenation. Attention ablation supported the contribution of cross-modal interaction to accuracy under stronger noise. At 15% noise, gating reduced the accuracy standard deviation by 56.9%, improving stability across training runs.
- In the scaled-jacket experiment, the AE detected 1183 of 1200 damaged samples, a detection rate of 98.58%. Among 900 undamaged samples, 83 false alarms gave a 9.22% false-alarm rate and 95.24% overall accuracy. Detection reached 100% for the short member and brace and 94.3% for the long member. Screening performance varied among member types, potentially reflecting differences in boundary constraints and modal-curvature sensitivity. The frozen classifier achieved 100% accuracy across seven states under a 5% low-frequency colored perturbation, using 1400 samples resampled from held-out impacts. It distinguished the prescribed member states and two short-member damage severities under this controlled laboratory test.
- The feature-extraction and classification pipeline required an average of 31.16 ms per window on the tested CPU, below the 200 ms window-update interval.

This study has several limitations. Evaluation was confined to a numerical model and a laboratory-scale jacket specimen, so offshore field validation remains necessary. Machined slots differ from progressive corrosion and fatigue cracks. Damage identification at welded joints, where fatigue cracks can initiate at weld toes [35], remains unvalidated. Dedicated

joint-level models and measurements are needed to establish applicability to these regions. The Gaussian beam tests and 5% low-frequency perturbation test do not reproduce the coupled offshore environment. The jacket was tested in air; the added disturbance did not represent immersion-induced added mass, hydrodynamic damping, wave-current loading, or underwater sensor behavior. Its frequency band and amplitude were prescribed rather than calibrated to measured sea conditions. The classification result is also limited to one specimen, limited acquisition batches, and one perturbation seed. Sampling-position errors, nonuniform spacing, and alternative layouts remain unresolved. Computational performance requires end-to-end verification on representative edge devices. Future work should combine submerged testing and environmental measurements with joint-level damage studies to establish applicability beyond the present laboratory conditions.

Author Contributions: Conceptualization, W.C.; methodology, W.C.; software, W.C.; validation, J.W. and X.Z.; formal analysis, W.C.; investigation, J.W. and X.Z.; resources, Z.F.; data curation, J.W. and X.Z.; writing—original draft preparation, W.C.; writing—review and editing, Z.F.; visualization, W.C.; supervision, Z.F.; project administration, Z.F.; funding acquisition, Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Dalian University of Technology through its Undergraduate Innovation and Entrepreneurship Training Program, grant number 20261014110374.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SHM	Structural Health Monitoring
MSCD	Mode-Shape Curvature Difference
MSD	Mode-Shape Difference
AE	Autoencoder
MLP	Multilayer Perceptron
1D CNN	One-Dimensional Convolutional Neural Network
MSE	Mean Squared Error
FDD	Frequency-Domain Decomposition
MNV3	MobileNetV3
SNV2	ShuffleNetV2
CrossATT	Cross-Attention
CMA	Cross-Modal Attention

References

1. Feng, S.; Song, L. Multi-Scale CNN for Health Monitoring of Jacket-Type Offshore Platforms with Multi-Head Attention Mechanism. *J. Mar. Sci. Eng.* **2025**, *13*, 1572. [CrossRef]
2. Wang, M.; Incecik, A. Structural Health Monitoring on Offshore Jacket Platforms Using a Novel Ensemble Deep Learning Model. *Ocean Eng.* **2024**, *301*, 117510. [CrossRef]
3. Cha, Y.-J.; Ali, R. Deep Learning-Based Structural Health Monitoring. *Autom. Constr.* **2024**, *161*, 105328. [CrossRef]
4. Jiang, C.; Chen, N.-Z. G-Twin: Graph Neural Network-Based Digital Twin for Real-Time and High-Fidelity Structural Health Monitoring for Offshore Wind Turbines. *Mar. Struct.* **2025**, *103*, 103813. [CrossRef]
5. Kenley, R.M.; Dodds, C.J. West Sole WE Platform: Detection of Damage by Structural Response Measurements. In *Proceedings of the Offshore Technology Conference*; OTC-3866-MS; Offshore Technology Conference: Houston, TX, USA, 1980; pp. 111–118. [CrossRef]
6. Killbourn, S.D. Detection of Jacket Member Severance Using Structural Instrumentation. In *Proceedings of the ASME 27th International Conference on Offshore Mechanics and Arctic Engineering, Estoril, Portugal, 15–20 June 2008*; Volume 2: Structures, Safety and Reliability; ASME: New York, NY, USA; pp. 379–387. [CrossRef]

7. Baybordi, S.; Esfandiari, A. Structural Damage Identification by Finite Element Model Updating Using Transmissibility Functions Data: Numerical and Experimental Study. *Mar. Struct.* **2026**, *106*, 103968. [CrossRef]
8. Lam, H.F.; Fu, Z.Y. Time-Domain Structural Model Updating Following the Bayesian Approach in the Absence of System Input Information. *Eng. Struct.* **2024**, *314*, 118321. [CrossRef]
9. Liu, J.; Li, Y. Physics-Encoded Interpretable Self-Supervised Learning for Structural Damage Identification. *Eng. Struct.* **2025**, *343*, 121045. [CrossRef]
10. Hejazitalab, H.; Taghikhany, T. Damage Localization and Quantification in Offshore Jacket Structures Using Signal Processing and Intelligent System. *Ocean Eng.* **2023**, *285*, 115325. [CrossRef]
11. Bao, X.; Liu, M. Damage Identification for Jacket Offshore Platforms Using Transformer Neural Networks and Random Decrement Technique. *Ocean Eng.* **2023**, *288*, 115973. [CrossRef]
12. Feng, H.; Leng, J. A Hankel Matrix Normalization and Improved BiLSTM-Integrated Damage Assessment Method for Jacket Platform Structures under Variable Excitations. *Ocean Eng.* **2026**, *363*, 126707. [CrossRef]
13. Abdelhak, M.; Prendergast, L.J. A Transfer Learning-Based Framework for Scour Detection around Monopile-Supported Offshore Wind Turbines with Limited Data. *Ocean Eng.* **2026**, *365*, 127315. [CrossRef]
14. Lv, J.; Diao, Y. Damage Identification of an Offshore Wind Turbine Support Structure Using VMD and Deep Transfer Learning. *Struct. Control Health Monit.* **2025**, *2025*, 1699730. [CrossRef]
15. Zhou, M.; Lai, Z. Structural Damage Classification under Varying Environmental Conditions and Unknown Classes via Open Set Domain Adaptation. *Mech. Syst. Signal Process.* **2024**, *218*, 111561. [CrossRef]
16. Li, S.; Cao, Y. Intelligent Framework for Unsupervised Damage Detection in Bridges Using Deep Convolutional Autoencoder with Wavelet Transmissibility Pattern Spectra. *Mech. Syst. Signal Process.* **2024**, *220*, 111653. [CrossRef]
17. Wei, A.; Guan, S. Damage Detection of Jacket Platforms through Improved Stacked Autoencoder and Softmax Classifier. *Ocean Eng.* **2024**, *306*, 118036. [CrossRef]
18. Bittner, M.; Grashorn, J. Unsupervised Anomaly Detection in Non-Linear Mechanical Systems for Structural Health Monitoring. *Acta Mech. Sin.* **2026**, *42*, 725585. [CrossRef]
19. Coraça, E.M.; Ferreira, J.V. An Unsupervised Structural Health Monitoring Framework Based on Variational Autoencoders and Hidden Markov Models. *Reliab. Eng. Syst. Saf.* **2023**, *231*, 109025. [CrossRef]
20. Leng, J.; Ma, J. An Adaptive Convolutional Neural Network Based on Transmissibility Grayscale Image for Online Identification of Offshore Platform Damage Pattern. *Mech. Syst. Signal Process.* **2024**, *221*, 111713. [CrossRef]
21. Nong, X.; Luo, X. Multimodal Deep Neural Network-Based Sensor Data Anomaly Diagnosis Method for Structural Health Monitoring. *Buildings* **2023**, *13*, 1976. [CrossRef]
22. Zhu, Y.; Wu, Y. Vision + X: A Survey on Multimodal Learning in the Light of Data. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 9102–9122. [CrossRef] [PubMed]
23. Xu, X.; Pan, T. A Novel Structural Damage Identification Method Based on Multi-Sensor Data Fusion and Multimodal Neural Networks. *Eng. Struct.* **2025**, *345*, 121512. [CrossRef]
24. Xu, X.; Lan, X. Multimodal Deep Learning with Integrated Automatic Labeling for Structural Damage Detection in High-Pile Wharves. *Ocean Eng.* **2025**, *340*, 122457. [CrossRef]
25. Cui, S.; Hoang, T. Adaptive Edge Intelligence for Rapid Structural Condition Assessment Using a Wireless Smart Sensor Network. *Eng. Struct.* **2025**, *326*, 119520. [CrossRef]
26. Dong, W.; Moan, T.; Gao, Z. Fatigue Reliability Analysis of the Jacket Support Structure for Offshore Wind Turbine Considering the Effect of Corrosion and Inspection. *Reliab. Eng. Syst. Saf.* **2012**, *106*, 11–27. [CrossRef]
27. Xiong, Q.; Kong, Q. Zero-Shot Knowledge Transfer for Seismic Damage Diagnosis through Multi-Channel 1D CNN Integrated with Autoencoder-Based Domain Adaptation. *Mech. Syst. Signal Process.* **2024**, *217*, 111535. [CrossRef]
28. Howard, A.; Sandler, M. Searching for MobileNetV3. In *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*; IEEE: New York, NY, USA, 2019; pp. 1314–1324.
29. Sandler, M.; Howard, A. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2018; pp. 4510–4520.
30. Vaswani, A.; Shazeer, N. Attention Is All You Need. In *Proceedings of the Advances in Neural Information Processing Systems 30*; Curran Associates, Inc.: Red Hook, NY, USA, 2017.
31. Song, D.; Qin, F. Few-Shot Learning with Label Smoothing and Metric Space Optimization for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5535722. [CrossRef]
32. Brincker, R.; Zhang, L. Modal Identification of Output-Only Systems Using Frequency Domain Decomposition. *Smart Mater. Struct.* **2001**, *10*, 441–445. [CrossRef]
33. Hendrycks, D.; Gimpel, K. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In Proceedings of the International Conference on Learning Representations, Toulon, France, 24–26 April 2018.

34. Ma, N.; Zhang, X. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. In *Computer Vision—ECCV*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 11218, pp. 122–138.
35. Ali, L.; Khan, S.; Bashmal, S.; Iqbal, N.; Dai, W.; Bai, Y. Fatigue Crack Monitoring of T-Type Joints in Steel Offshore Oil and Gas Jacket Platform. *Sensors* **2021**, *21*, 3294. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.