mathematics

MDPI

Article

Cross-Modal Degradation Rivalry for Self-Supervised Structural Fatigue Health Monitoring

Tianbao Nie [1], Yu Yang [2,3] and Xiang Li [1,*]

[1] Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an 710049, China
[2] National Key Laboratory of Strength and Structural Integrity, Xi'an 710065, China
[3] Aircraft Strength Research Institute of China, Xi'an 710065, China
* Correspondence: lixiang@xjtu.edu.cn

Abstract

Fatigue health monitoring of engineering structures requires continuous degradation assessment, yet ground-truth health labels are unavailable during run-to-failure tests. Existing self-supervised approaches rely on monotonic degradation assumptions that are violated by the structured non-monotonic behaviour of acoustic emission signals during fatigue. A self-supervised framework called Cross-Modal Degradation Rivalry (CMDR) is proposed, which introduces the Modal Rivalry Index (MRI) as a directional measure of cross-modal predictability between heterogeneous sensor modalities. CMDR comprises a label-free representation-learning stage trained via the Cross-Modal Prediction Asymmetry (CMPA) pretext task, followed by a lightweight supervised stage that maps MRI features to scalar health indicators (HIs) using normalised lifecycle labels. The MRI is conceptually related, under the stated assumptions only loosely met in practice, to the Transfer Entropy difference between sensor latent channels. Experiments on a structural fatigue dataset with seven specimens under two loading conditions demonstrate that CMDR achieves competitive trendability and prognosability, as well as the lowest remaining useful life (RUL) error in three of four scenarios. RUL evaluations are additionally repeated under a fully online estimator that uses only training specimens. A strictly inductive ablation that re-pre-trains the self-supervised stage within each leave-one-specimen-out fold confirms a bounded transductive-vs-inductive gap, and CMDR remains the best against three further self-supervised baselines on the within-condition and mixed-condition scenarios. Ablation studies confirm the necessity of directional asymmetry, bottleneck architecture, and momentum-updated target encoders.

Keywords: structural health monitoring; self-supervised learning; health indicator construction; cross-modal prediction; fatigue degradation; remaining useful life

MSC: 68T07

Academic Editor: Zdeněk Kala

Received: 28 May 2026
Revised: 17 June 2026
Accepted: 20 June 2026
Published: 23 June 2026

1. Introduction

Structural fatigue is one of the most common failure mechanisms in civil infrastructure, aerospace components, and mechanical systems [1–3], and health monitoring across the full lifecycle—from initial loading through crack initiation, stable propagation, and final fracture—is essential for condition-based maintenance and the prevention of catastrophic failures [4]. Modern monitoring systems deploy multiple heterogeneous sensor types,

including acoustic emission (AE), strain gauges, and fibre optic sensors, each capturing complementary physical aspects of the degradation process [5].

A fundamental obstacle is the absence of ground-truth health labels. Run-to-failure experiments produce continuous multi-sensor streams, but the only available supervisory signal is the failure time itself, making supervised health indicator (HI) construction infeasible in practice [6–8]. This has motivated growing interest in self-supervised and unsupervised HI construction, where representations are learned from the data itself without explicit degradation labels [9,10]. They are typically built on deep neural networks such as convolutional neural networks, recurrent neural networks, long short-term memory, and so forth [11–13].

Recent self-supervised methods for HI construction rely predominantly on temporal ordering as the pretext task, assuming monotonic degradation [9]. This holds for bearings and turbofan engines but is fundamentally violated in AE-based fatigue monitoring. AE signals exhibit a canonical three-phase activity pattern [14]: (i) high activity during initial micro-yielding, (ii) a quiet period during stable microcrack incubation, and (iii) renewed intense activity during macro-crack propagation. This canonical pattern is best read as one prototypical instance of a broader family of structured non-monotonic trajectories. The quantitative phase analysis in the experimental study shows that only a subset of specimens exhibits the textbook three-phase ordering, the remainder displaying alternative non-monotonic structures. What breaks monotonicity-based methods is therefore not the universality of the three-phase template but the non-monotonic nature of AE in fatigue. Temporal ordering therefore does not correspond to degradation ordering for AE features.

Meanwhile, strain signals follow a quasi-monotonic trajectory, and fibre optic sensors measure local deformation at discrete locations. These heterogeneous modalities respond to the same degradation process at different rates and through different mechanisms. Early-life AE events reveal incipient damage not yet visible in strain. During the quiet period, both are quiescent and mutually predictable. Late-life strain accelerates and carries information about imminent failure that discrete AE bursts cannot fully capture.

This evolving directional asymmetry is illustrated in Figure 1 (an idealised trajectory, while empirically observed MRI trajectories vary across specimens). The central hypothesis is that the temporal trajectory of this asymmetry constitutes a label-free degradation indicator requiring no monotonicity assumption and no external supervision for the representation-learning stage. A small downstream regressor maps MRI features to a scalar HI using lifecycle-ratio supervision.

The proposed approach differs from existing information-theoretic methods for damage detection. Mutual information (MI) [15] is symmetric and provides no directional information. Granger causality methods [16] target anomaly detection rather than continuous degradation tracking. Transfer Entropy [17] captures directional information flow but requires density estimation that scales poorly with dimensionality. The proposed approach instead computes an asymmetry of prediction residuals in a learned latent space, which, under given assumptions, behaves as a proxy for the Transfer Entropy difference.

Based on this observation, the Cross-Modal Degradation Rivalry (CMDR) framework introduces the Modal Rivalry Index (MRI)—a scalar measure of the directional asymmetry in cross-modal predictability between sensor pairs. MRI is computed from bottleneck cross-modal predictors trained via the Cross-Modal Prediction Asymmetry (CMPA) pretext task, and the MRI trajectory over the fatigue lifecycle serves directly as the health indicator.

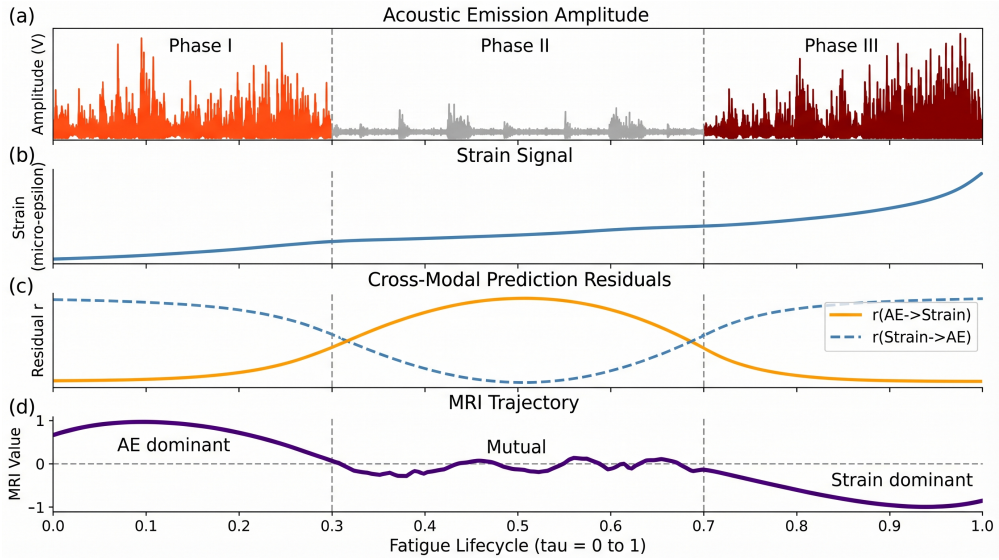

Figure 1. Conceptual illustration of the three-phase fatigue behaviour and the resulting Cross-Modal Prediction Asymmetry. (**a**) Non-monotonic three-phase AE amplitude. (**b**) Quasi-monotonic strain accumulation. (**c**) Direction-dependent evolution of cross-modal prediction residuals. (**d**) The Modal Rivalry Index (MRI), the directional asymmetry of prediction residuals, traces a characteristic trajectory through the three phases.

The contributions of this work are summarised as follows:

1. The Modal Rivalry Index (MRI) is introduced as a new concept for degradation monitoring. MRI measures the directional asymmetry of cross-modal predictability and, under the assumptions of bottleneck-predictor optimality and Gaussian latent residuals, behaves as a proxy for the Transfer Entropy difference [17,18] between sensor channels.
2. The Cross-Modal Prediction Asymmetry (CMPA) pretext task is proposed for self-supervised representation learning. CMPA trains bottleneck cross-modal predictors whose residual asymmetry serves as the health indicator, requiring no monotonicity assumption and no degradation labels.
3. The well-known three-phase AE behaviour during fatigue, which is an obstacle for monotonicity-based methods, is demonstrated to become a powerful feature when viewed through the lens of inter-modal prediction dynamics.
4. Comprehensive experiments on a structural fatigue dataset with seven specimens under two loading conditions across four evaluation scenarios demonstrate that CMDR achieves competitive or superior trendability, prognosability, and remaining useful life (RUL) prediction compared to six baseline methods. Statistical significance is assessed via paired Wilcoxon signed-rank testing, and an online RUL estimator without access to the test specimen's true lifetime is additionally evaluated. Cohen's d affects sizes with 95% bootstrap confidence intervals, per-seed performance distributions, a strictly inductive per-fold re-pre-training ablation, three additional modern SSL baselines (BYOL-style, masked reconstruction, transformer contrastive), and an adversarial synthetic study further substantiate the conclusions.

2. Related Work

2.1. Self-Supervised Learning for Health Indicator Construction

Self-supervised learning (SSL) has emerged as a promising approach for constructing health indicators without explicit degradation labels [19]. Rombach et al. [9] used operational time as a proxy for degradation ordering in a contrastive framework built on

popular deep learning architectures [20,21]. The approach performs well on bearings and milling but assumes monotonic degradation, making it unsuitable for AE-based fatigue monitoring. Akrim et al. [10] applied SSL to fatigue prognostics with strain sensors, and a recent survey on few-shot RUL prediction [22] identified SSL as a promising but underexplored direction. Broader SSL paradigms recently adopted in time-series prognostics include BYOL-style momentum self-prediction [23], masked time-series reconstruction (the MAE family [24] and transformer masked-imputation for multivariate series [25]), and temporal–contextual contrastive learning exemplified by TS-TCC. CMDR re-uses engineering ingredients popularised by these works, but its core mechanism is orthogonal: those paradigms learn representations that are *stable* across views or *reconstruct* input content, while CMDR explicitly constructs a directional *asymmetry* between modalities and uses its temporal evolution as the health indicator, and a head-to-head comparison is reported in the experimental study. Beyond fatigue, Pakzad and Masoodi [26] recently proposed a variational-autoencoder hybrid for early bridge-damage detection based on reconstruction-based anomaly scoring.

2.2. Cross-Modal Learning for Time Series

CroSSL [27] introduced cross-modal self-supervised learning via latent masking, where one modality is masked and reconstructed from the other. Reconstruction is symmetric and does not track the temporal evolution of cross-modal relationships. Shared-specific feature modelling addresses missing modalities for general multimodal fusion [28,29], and the multimodal-ML literature provides theoretical foundations [30,31], but direct application to degradation monitoring requires domain-specific design. SHM fusion reviews [5,32] reveal that most approaches combine sensor information at a single time point without modelling how inter-sensor relationships change over the lifecycle.

2.3. Information-Theoretic Approaches to Damage Detection

Transfer Entropy (TE), introduced by Schreiber [17], measures directional information flow between processes, extending MI with directionality. Kraskov et al. [18] introduced the k-NN MI estimator underpinning many TE estimators on continuous-valued series, and Barnett et al. [33] proved the Granger–TE equivalence for Gaussian variables. GCAD [16] used gradient-based Granger causality in deep predictors for multivariate anomaly detection, and Tezcan and Marin-Artieda [15] used MI between sensors to detect damage in nonlinear systems. These works establish that directional information flow changes with structural condition. The present work differs by operationalising directional asymmetry as a continuous HI trajectory rather than a binary anomaly indicator, with the information-theoretic interpretation conditional on the stated modelling assumptions.

2.4. Acoustic Emission in Fatigue Monitoring

Roberts and Talebzadeh [14] documented the three-stage AE pattern during fatigue crack propagation in steel: high initial activity, a quiescent intermediate period, and renewed activity during rapid crack growth. Chai et al. [34] applied ML to AE-based fatigue crack prediction, and Dong et al. [35] combined Paris' law with physics-informed neural stochastic differential equations and AE data. These works treat AE as a single-modality tool and the three-phase behaviour as a characteristic to be modelled. The present work instead exploits it as a feature, creating distinctive Cross-Modal Prediction Asymmetry patterns when combined with strain.

3. Methodology

3.1. Problem Formulation

Consider a set of N specimens $\{X^{(1)}, \ldots, X^{(N)}\}$ undergoing fatigue loading until failure. For specimen i, the multi-modal observation at time step t is denoted as $\mathbf{x}_t^{(i)} = \{\mathbf{a}_t, s_t, \mathbf{f}_t\}$, where $\mathbf{a}_t \in \mathbb{R}^{25}$ represents the 25 time- and frequency-domain AE statistical features extracted from each 0.1 s segment, $s_t \in \mathbb{R}^1$ represents strain amplitude, and $\mathbf{f}_t \in \mathbb{R}^{d_f}$ represents fibre optic measurements with $d_f \in \{0, 1, 5\}$, depending on the sensor configuration of the specimen. The available modality set for specimen i is $\mathcal{M}^{(i)} \subseteq \{A, S, F\}$, where A and S are always available and F is optional.

The lifecycle ratio $\tau_t^{(i)} = t/T^{(i)} \in [0, 1]$ is the normalised time, where $T^{(i)}$ is the total lifetime. It is used only as supervision for the downstream HI fine-tuning stage and is never accessed during self-supervised pre-training. The objective is to construct a health indicator $\mathrm{HI}(t) \in [0, 1]$ that increases monotonically toward failure and enables RUL prediction $\widehat{\mathrm{RUL}}(t) = (1 - \mathrm{HI}(t)) \times \hat{T}$, where $\hat{T}$ is a lifetime estimator. For online prediction, where the test specimen's lifetime is unavailable, $\hat{T}$ is restricted to functions of the training specimens.

3.2. Framework Overview and Three-Stage Pipeline

The CMDR pipeline comprises three stages with explicit label-availability:

- Stage 1 (label-free pre-training). The 1D-CNN encoders, EMA target encoders, and bottleneck cross-modal predictors are trained via CMPA using *no* degradation labels.
- Stage 2 (MRI extraction). The trained network is run in inference mode to produce per-specimen MRI trajectories.
- Stage 3 (supervised HI/RUL fine-tuning). A small GRU head is trained on training specimens only to map MRI features to the lifecycle ratio τ_t, with $\hat{T}$ likewise built from training specimens only.

CMDR is thus a self-supervised representation-learning framework with supervised HI fine-tuning rather than a fully label-free pipeline (Figures 2 and 3).

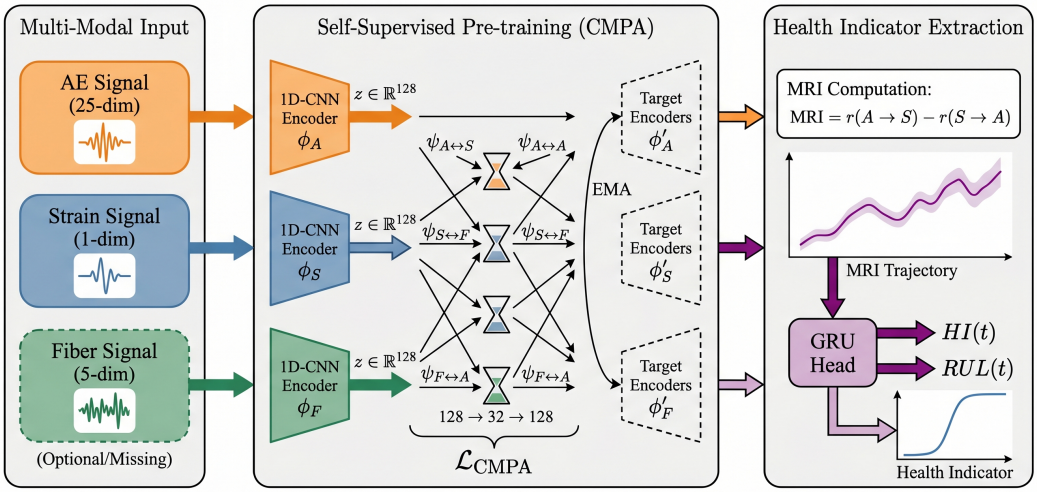

Figure 2. Overview of the CMDR framework. Modality-specific 1D-CNN encoders map each sensor signal to a shared latent space. Bottleneck cross-modal predictors estimate each modality from every other, with EMA target encoders providing stable prediction targets. The Modal Rivalry Index is computed as the directional asymmetry of prediction residuals.

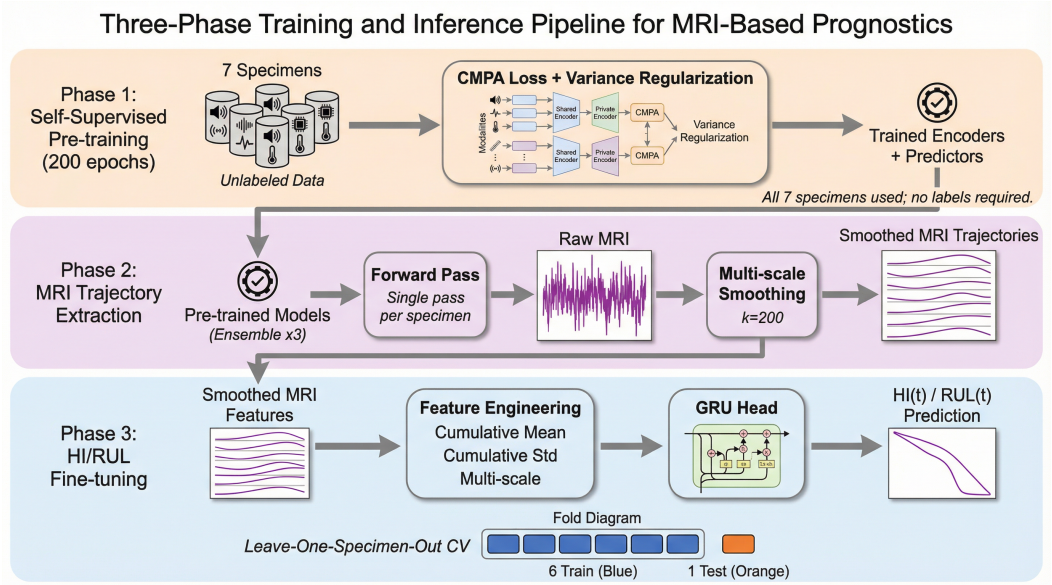

Figure 3. The three-stage training pipeline: (1) label-free self-supervised pre-training via CMPA on all specimens, (2) MRI trajectory extraction with ensemble averaging and smoothing, and (3) HI/RUL fine-tuning via leave-one-specimen-out cross-validation with lifecycle-ratio supervision on training specimens only.

3.3. Modality-Specific Encoders

Each modality is encoded by a dedicated 1D-CNN encoder into a shared d-dimensional latent space ($d = 128$). For a window of $2w + 1$ time steps centred at t (with half-window $w = 25$), the encoders produce:

$$\mathbf{z}_t^A = \phi_A(\mathbf{a}_{t-w:t+w}) \in \mathbb{R}^d, \tag{1}$$

$$\mathbf{z}_t^S = \phi_S(s_{t-w:t+w}) \in \mathbb{R}^d, \tag{2}$$

$$\mathbf{z}_t^F = \phi_F(\mathbf{f}_{t-w:t+w}) \in \mathbb{R}^d. \tag{3}$$

The encoder window is centred at t and thus uses both past and future samples, a non-causal design appropriate for the offline post hoc analysis constituting the primary evaluation. The causal left-aligned variant required for online deployment is evaluated in the online RUL experiments.

Each encoder uses three 1D convolutional layers with kernel sizes 7, 5, and 3, followed by batch normalization, GELU activation, global average pooling, a linear projection to $\mathbb{R}^d$, and layer normalization. Channel dimensions are 25→64→64→128 for AE, 1→32→64→128 for Strain, and 5→32→64→128 for Fibre. Modality-specific encoders accommodate the differing input dimensions and physical semantics while projecting into a common 128-dimensional latent space.

The window size $w = 25$ (5.0 s at 0.1 s sampling, or 50 loading cycles at 10 Hz) matches the characteristic timescale of AE events, capturing local damage activity without spanning multiple fatigue phases.

3.4. Bottleneck Cross-Modal Predictors

For each ordered pair of modalities (m_1, m_2) with $m_1 \neq m_2$, a bottleneck predictor network estimates the target modality representation from the source modality representation:

$$\hat{\mathbf{z}}_t^{m_1 \to m_2} = \psi_{m_1 \to m_2}(\mathbf{z}_t^{m_1}). \tag{4}$$

Each predictor $\psi_{m_1 \to m_2}$ is a two-layer MLP with a bottleneck: $d \to d/4 \to d$ (i.e., 128→32→128), with layer normalization and GELU activation at the bottleneck layer. The detailed architecture is shown in Figure 4.

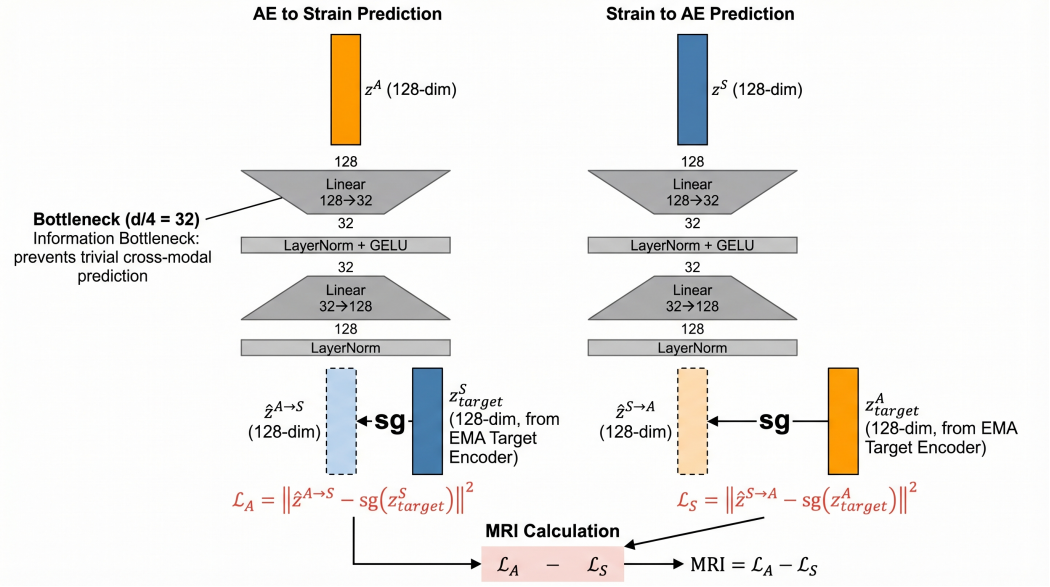

Figure 4. Architecture of the bottleneck cross-modal predictor. The bottleneck layer (dimension $d/4 = 32$) constrains inter-modal information bandwidth so that residuals reflect genuine information gaps rather than capacity limitations [36].

The bottleneck is motivated by the Information Bottleneck principle [36]: without it, sufficiently large predictors could memorise near-perfect mappings, driving MRI uniformly near zero, whereas compression forces residuals to reflect genuine inter-modal information gaps. The predictors are directional: $\psi_{A\to S}$ and $\psi_{S\to A}$ are independent networks because the two directions capture different information flow. Three modalities yield six directional predictors, and two modalities yield two.

3.5. EMA Target Encoders

To prevent representation collapse during self-supervised pre-training, target representations are produced by exponential moving average (EMA) copies of the online encoders, following the BYOL [23] design:

$$\theta_{\text{tgt}} \leftarrow \beta \cdot \theta_{\text{tgt}} + (1 - \beta) \cdot \theta_{\text{online}}, \tag{5}$$

where β is the momentum coefficient, initialised at 0.996 and linearly increased toward 0.999 during training. The targets are computed with stop-gradient:

$$\mathbf{z}^m_{t,\text{tgt}} = \mathrm{sg}(\phi_{m,\text{tgt}}(\mathbf{x}^m_{t-w:t+w})), \tag{6}$$

where $\mathrm{sg}(\cdot)$ denotes the stop-gradient operation. Gradients flow only through the online encoders and predictors, preventing the trivial solution where all representations collapse to a constant.

3.6. Modal Rivalry Index

The prediction residual for direction $m_1 \to m_2$ at time t is defined as:

$$r_t^{m_1\to m_2} = \|\hat{\mathbf{z}}_t^{m_1\to m_2} - \mathbf{z}_t^{m_2}\|_2^2. \tag{7}$$

The Modal Rivalry Index for a modality pair (m_1, m_2) is the directional asymmetry of prediction residuals:

$$\mathrm{MRI}_t^{(m_1,m_2)} = r_t^{m_1 \to m_2} - r_t^{m_2 \to m_1}. \tag{8}$$

When $\mathrm{MRI}_t^{(A,S)} > 0$, AE predicts Strain less accurately than Strain predicts AE, interpreted as AE carrying unique information not recoverable from Strain within the bottleneck budget (e.g., discrete micro-damage events during Phase 1). When $\mathrm{MRI}_t^{(A,S)} < 0$, the reverse holds—Strain carries information that AE cannot predict (e.g., accelerating macro-strain during Phase 3). The temporal trajectory $\{\mathrm{MRI}_t^{(A,S)}\}_{t=1}^{T}$ is expected to trace a characteristic curve reflecting the three fatigue phases.

Relation to the Transfer Entropy difference. The connection between MRI and Transfer Entropy [17] relies on three assumptions: (A1) $\psi_{m_1 \to m_2}$ approximates the conditional mean $\mathbb{E}[\mathbf{z}^{m_2} \mid \mathbf{z}^{m_1}]$ within its capacity; (A2) the residual is approximately Gaussian with covariance proportional to identity; and (A3) the encoders are stationary at the local window timescale. Under (A1)–(A3), the squared ℓ_2 residual norm is affine in the conditional differential entropy [18]:

$$r^{m_1 \to m_2} \propto H(\mathbf{Z}^{m_2} \mid \mathbf{Z}^{m_1}) + \text{const.} \tag{9}$$

The constants cancel in the difference, so

$$\mathrm{MRI}^{(A,S)} \propto H(\mathbf{Z}^S \mid \mathbf{Z}^A) - H(\mathbf{Z}^A \mid \mathbf{Z}^S), \tag{10}$$

which is the latent-space analogue of the TE difference $\mathrm{TE}(A \to S) - \mathrm{TE}(S \to A)$. The proportionality is not strict because MRI uses same-window latents, the bottleneck restricts predictor capacity, and Gaussianity holds only approximately. MRI is thus treated as a cross-modal predictability-asymmetry measure conceptually aligned with the TE difference, with an empirical correlation to a k-NN TE-difference estimator reported in the experimental study.

Four formal properties follow. First, informationally symmetric modalities yield MRI near zero (the quiet Phase 2). Second, a sign change indicates a reversal of the dominant information-flow direction. Third, $|\mathrm{MRI}_t|$ is bounded by the maximum residual, constrained by the variance regularisation. Fourth, statistically independent modalities yield MRI near zero with high individual residuals, distinguishable from the low-residual Phase 2 case.

3.7. Self-Supervised Pre-Training Loss

The CMPA pre-training objective minimises all cross-modal prediction errors simultaneously:

$$\mathcal{L}_{\mathrm{CMPA}} = \frac{1}{|\mathcal{P}|} \sum_{(m_1,m_2)\in\mathcal{P}} \mathbb{E}_t \left[\| \hat{\mathbf{z}}_t^{m_1 \to m_2} - \mathrm{sg}(\mathbf{z}_{t,\mathrm{tgt}}^{m_2}) \|_2^2 \right], \tag{11}$$

where $\mathcal{P}$ is the set of all ordered modality pairs available for the current specimen.

A variance regularisation term, following VICReg [37], prevents representation collapse:

$$\mathcal{L}_{\mathrm{var}} = \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} \mathrm{mean}\left(\max\left(0, \gamma - \sqrt{\mathrm{Var}[\mathbf{z}_t^m] + \epsilon}\right)\right), \tag{12}$$

where $\gamma = 1.0$ is the target standard deviation and $\epsilon = 10^{-4}$ is a small constant. The total pre-training loss is:

$$\mathcal{L}_{\mathrm{pretrain}} = \mathcal{L}_{\mathrm{CMPA}} + \lambda_{\mathrm{var}} \mathcal{L}_{\mathrm{var}}, \tag{13}$$

with $\lambda_{\text{var}} = 1.0$.

Pre-training runs for 200 epochs on all N specimens using AdamW [38] with cosine annealing [39] and 10-epoch warmup. Only available modality pairs contribute to the loss, naturally handling the missing fibre modality. The implications of pre-training on all specimens versus per-fold re-pre-training are discussed later in this section.

3.8. MRI Trajectory Extraction and Ensemble

After pre-training, the MRI trajectory is computed for each specimen by a forward pass through the trained encoders and predictors, using Equation (8). To reduce sensitivity to initialisation, three models pre-trained with different random seeds are ensembled by element-wise averaging of their MRI trajectories, which are then smoothed with a centred moving-average kernel of size $k = 200$. A causal (left-aligned) smoother is additionally evaluated in the online RUL experiments.

For HI mapping, three features are engineered from the smoothed MRI trajectory: the smoothed MRI value, its cumulative running mean, and its cumulative running standard deviation, capturing both the instantaneous cross-modal prediction state and its historical evolution. A dedicated ablation confirms that the MRI itself carries the majority of the predictive signal.

3.9. HI and RUL Fine-Tuning

A single-layer Gated Recurrent Unit (GRU) [40] with 64 hidden units maps the MRI feature sequence to a scalar health indicator:

$$\text{HI}(t) = \sigma(\text{Linear}(\text{GRU}(\mathbf{H}_{t-W:t}))), \tag{14}$$

where $\mathbf{H}_t$ is the three-dimensional MRI feature vector at time t, W is the input window size, and $\sigma(\cdot)$ is the sigmoid activation. The GRU is trained with MSE loss against the lifecycle ratio τ_t of the training specimens. In the primary experiments $\hat{T} = T$, the true lifetime of the test specimen, while the online evaluation replaces $\hat{T}$ with three estimators depending only on training specimens.

Hyperparameter selection uses nested leave-one-specimen-out cross-validation (LOO-CV): for each outer fold, an inner LOO-CV loop over the remaining $N - 1$ specimens selects the best learning rate $\in \{10^{-4}, 5 \times 10^{-4}, 10^{-3}\}$ and window size $W \in \{25, 50, 100\}$, after which the final GRU is trained on all $N - 1$ training specimens and evaluated on the held-out specimen. This nested design prevents leakage from hyperparameter selection to test evaluation.

3.10. Pre-Training Scope and the LOO Evaluation Protocol

The self-supervised pre-training in Stage 1 uses raw multi-sensor signals from *all* seven specimens (including each fold's held-out specimen) without any labels. LOO-CV is enforced only at Stage 3, where the test specimen is unseen by the supervised GRU head—a *transductive* self-supervised setting. All baselines (PCA, Autoencoder, Contrastive-HI, DegCLF, CNN-GRU, LSTM) follow the same access policy, matching the standard fairness convention in SSL-PHM benchmarks [9]. To quantify the optimism this convention may introduce, a strictly inductive variant is additionally reported, in which Stage 1 is re-trained from scratch on the six training specimens of every fold, so that the held-out specimen contributes no samples, labelled or unlabelled, to that fold's encoders.

The CMDR model has ∼245 k parameters (180 k encoders, 52 k predictors, and 13 k GRU head, plus 180 k frozen EMA target encoders), a lightweight design ensuring that residuals reflect genuine information gaps rather than capacity effects.

The pairwise architecture naturally handles missing modalities: only available pairs contribute to the CMPA loss and to HI extraction. Primary evaluation uses AE–Strain only across all specimens, for consistent feature dimensionality and maximum LOO-CV statistical power. A three-modal experiment on Condition 1 specimens is reported in the experimental study.

4. Experimental Study

4.1. Dataset and Experimental Setup

Experiments are conducted on a structural fatigue full-lifecycle dataset from laboratory testing of structural steel specimens (Figure 5). The sensor arrangement is shown in Figure 6. AE signals were acquired at 1000 kHz while strain and fibre optic signals were sampled at 10 Hz. Every 100,000 AE points (corresponding to 0.1 s) form one analysis unit, from which 25 time- and frequency-domain features are extracted (mean square, RMS, peak, peak-to-peak, min, max, std, variance, mean/root-mean absolute value, skewness, kurtosis, waveform/impulse/peak/margin/kurtosis/skewness factors, spectral energy, mean squared frequency, RMS frequency, spectral variance/kurtosis/skewness). The 0.1 s segments align all three modalities on a unified time scale.

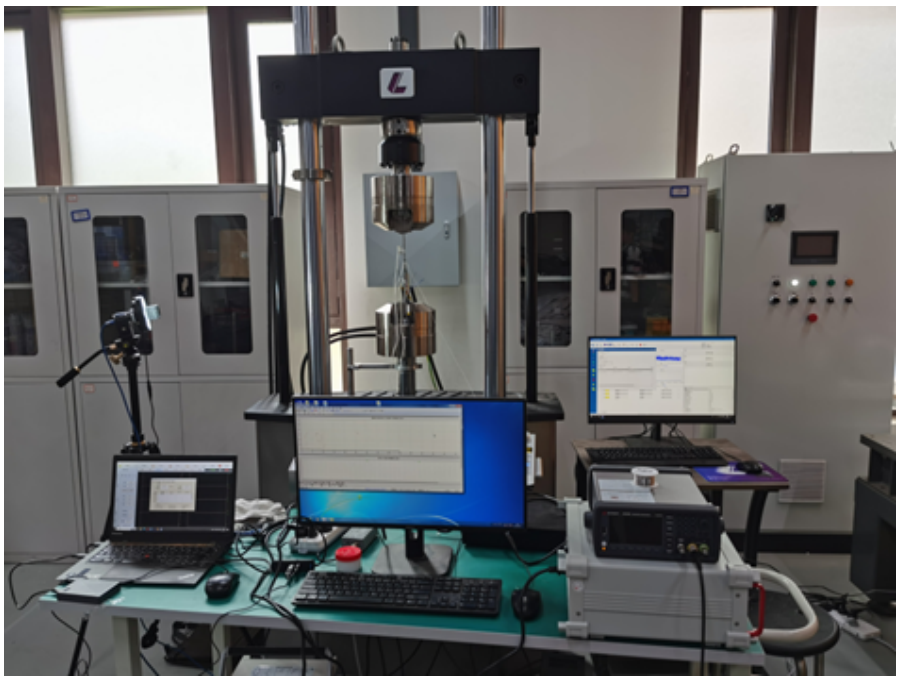

Figure 5. Overall structure of the fatigue tensile test bench.

Per-specimen preprocessing comprises (i) AE feature extraction yielding $\mathbf{a}_t \in \mathbb{R}^{25}$, (ii) $5 \cdot \mathrm{MAD}$ outlier rejection on strain/fibre in a 200-sample sliding window with linear interpolation, (iii) per-specimen z-score normalisation on the training portion only, and (iv) timestamp alignment on the 0.1 s grid. Steps (ii)–(iii) are recomputed within each LOO fold to prevent test-specimen leakage.

The dataset contains seven specimens subjected to sinusoidal cyclic loading until failure under two loading conditions:

- Condition 1 (high load): 10 Hz, 10 kN peak/1 kN valley, stress ratio 0.1. Four specimens (016, 017, 018, 019) with lifetimes of 30 k–46 k time steps (0.1 s per step).
- Condition 2 (low load): 10 Hz, 8 kN peak/0.8 kN valley, stress ratio 0.1. Three specimens (022, 024, 025) with lifetimes of 33 k–49 k time steps.

Condition 1 specimens have 1–5 fibre sensors, while Condition 2 specimens have none, providing a realistic testbed for variable modality availability.

Four evaluation scenarios assess different aspects of generalisation:

- S1: Within-Condition 1 LOO-CV (4 folds);
- S2: Within-Condition 2 LOO-CV (3 folds);
- S3: Cross-condition transfer (train Cond. 1 $\rightarrow$ test Cond. 2);
- S4: Mixed LOO-CV across all 7 specimens (7 folds).

The reverse transfer (Cond. 2 $\rightarrow$ Cond. 1) is excluded because three fibre-free Condition 2 specimens provide insufficient training data for the richer Condition 1 configuration.

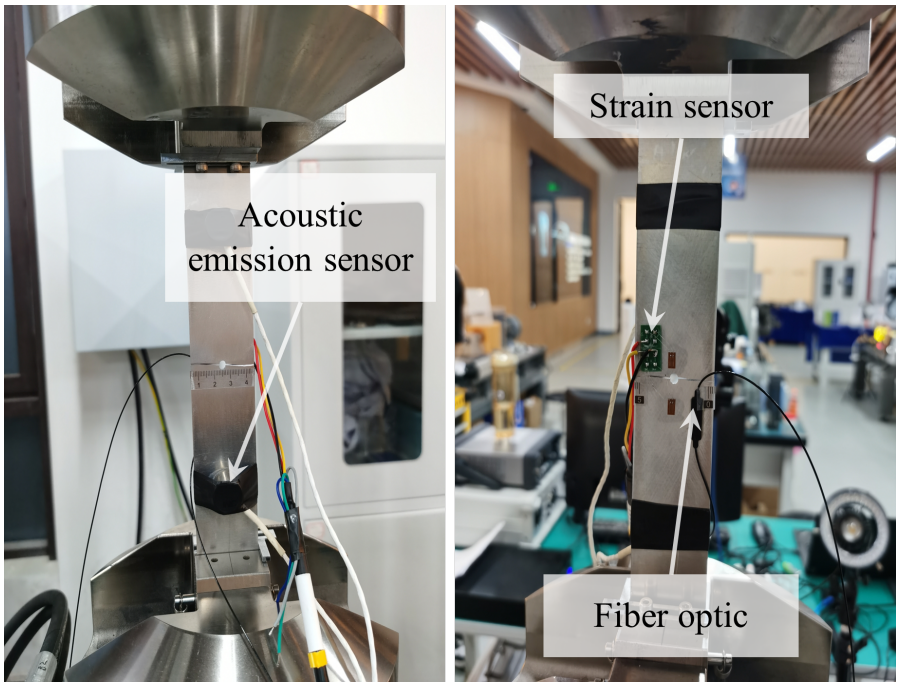

Figure 6. Arrangement of the test piece and sensors.

4.2. Evaluation Metrics

Let $\{\mathrm{HI}_t\}_{t=1}^{T}$ be an HI trajectory, $\tau_t = t/T$ the lifecycle ratio, and $\mathrm{HI}_T^{(i)}$ the end-of-life HI value of specimen i over K specimens.

- Trendability: absolute Spearman rank correlation between HI and τ, $\mathrm{Tre} = \left|\rho_{\mathrm{Spearman}}(\{\mathrm{HI}_t\}, \{\tau_t\})\right| \in [0, 1]$.
- Prognosability (standard PHM formulation):

$$\mathrm{Pro} = 1 - \frac{\mathrm{std}_i(\mathrm{HI}_T^{(i)})}{\frac{1}{K}\sum_{i=1}^{K} |\mathrm{HI}_T^{(i)} - \mathrm{HI}_0^{(i)}|}. \tag{15}$$

  Negative values indicate that the end-of-life spread across specimens exceeds the within-specimen HI excursion. The extreme -460.2 for PCA on S2 arises from a near-zero HI range and is reported as-is to expose degenerate baselines.
- RUL RMSE: $\sqrt{\frac{1}{T}\sum_t (\widehat{\mathrm{RUL}}_t - \mathrm{RUL}_t)^2}$.
- Monotonicity: $\mathrm{Mon} = |n_+ - n_-|/(T-1)$, $n_\pm = \#\{t : \mathrm{HI}_{t+1} - \mathrm{HI}_t \gtrless 0\}$.

4.3. Baseline Methods

Six baselines spanning unsupervised, self-supervised, and supervised approaches are compared:

- PCA: first principal component of all available sensor features as HI [41].
- Autoencoder (AE-Recon): feedforward autoencoder trained on early-life data, with reconstruction error as HI.
- CNN-GRU: supervised 1D-CNN + GRU model trained directly on lifecycle ratio.
- Contrastive-HI: temporal-ordering contrastive learning following Rombach et al. [9], adapted to multi-sensor input by concatenating sensor features before encoding.
- DegCLF: degradation-ordered contrastive learning with multi-modal fusion, using a triplet loss with temporal ordering as the degradation proxy.
- LSTM: a simple LSTM trained on raw sensor features to predict lifecycle ratio.

All neural baselines use comparable sizes (150 k–300 k parameters) and the same optimiser settings as CMDR: AdamW (lr 5×10^{-4}, wd 10^{-4}, batch 256, 200 epochs, 10-epoch warm-up) on the same z-score-normalised features and sliding windows. The autoencoder uses a 256-32-256 stack. Contrastive baselines use NT-Xent (temperature 0.1, 32-dim projection). Early stopping uses a 10% held-out split with patience 20.

4.4. Main Results

4.4.1. Trendability

Table 1 reports trendability across scenarios. CMDR is best in three of four (S1 0.491, S2 0.811, S4 0.787). In S3, CNN-GRU (0.566) is slightly higher than CMDR (0.504), which remains competitive. Figure 7 visualises the comparison.

Table 1. Trendability comparison (mean ± std, 5 seeds). Best results in bold. Higher is better.

Method	S1	S2	S3	S4
PCA	0.163	0.178	0.175	0.142
AE-Recon	0.473 ± 0.008	0.329 ± 0.072	0.355 ± 0.019	0.344 ± 0.025
CNN-GRU	0.406 ± 0.051	0.294 ± 0.070	**0.566** ± 0.038	0.380 ± 0.084
Contr-HI	0.184 ± 0.008	0.386 ± 0.033	0.251 ± 0.021	0.251 ± 0.029
DegCLF	0.469 ± 0.057	0.598 ± 0.034	0.479 ± 0.022	0.474 ± 0.023
LSTM	0.434 ± 0.063	0.352 ± 0.094	0.442 ± 0.031	0.453 ± 0.027
CMDR	**0.491** ± 0.168	**0.811** ± 0.052	0.504 ± 0.205	**0.787** ± 0.120

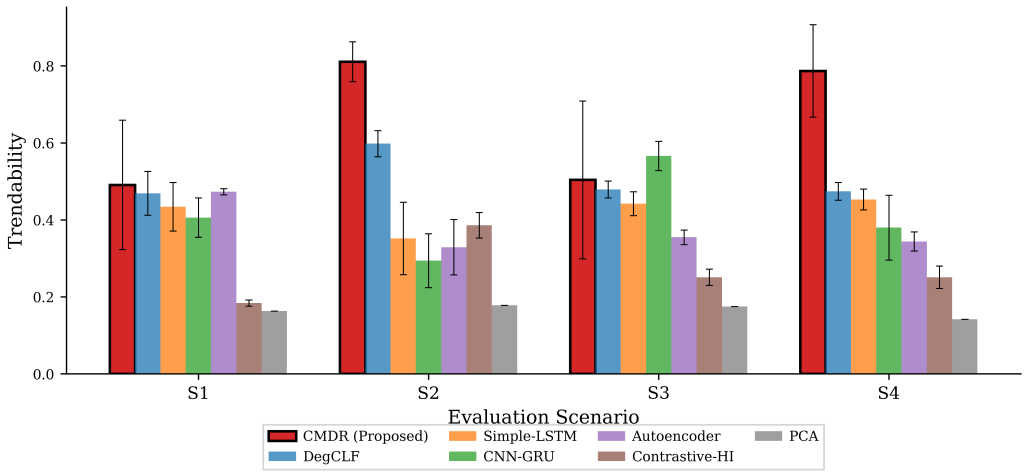

Figure 7. Trendability comparison across four evaluation scenarios.

The S2 and S4 gaps are substantial: CMDR 0.811 vs. DegCLF 0.598 (+36%) on S2, and 0.787 vs. 0.474 (+66%) on S4, showing the AE–Strain pair alone is highly informative.

4.4.2. Prognosability

Table 2 shows that CMDR is best in three of four scenarios (S2 0.907, S3 0.695, S4 0.761) with large margins. In S1, LSTM (0.604) outperforms CMDR (0.410), reflecting the difficulty of four-fold LOO on Condition 1. The prognosability comparison across four evaluation scenarios are presented in Figure 8.

Table 2. Prognosability comparison (mean ± std, 5 seeds). Best results in bold. Higher is better.

Method	S1	S2	S3	S4
PCA	−1.560	−460.2	−0.623	−0.984
AE-Recon	−0.345 ± 0.091	−0.543 ± 0.640	0.118 ± 0.086	−0.397 ± 0.105
CNN-GRU	0.386 ± 0.274	0.452 ± 0.167	0.433 ± 0.208	−0.080 ± 0.130
Contr-HI	−1.038 ± 0.595	0.273 ± 0.129	0.258 ± 0.358	−0.346 ± 0.122
DegCLF	0.444 ± 0.031	0.070 ± 0.156	0.468 ± 0.157	0.087 ± 0.024
LSTM	**0.604** ± 0.266	0.395 ± 0.301	0.390 ± 0.035	0.379 ± 0.067
CMDR	0.410 ± 0.228	**0.907** ± 0.033	**0.695** ± 0.220	**0.761** ± 0.099

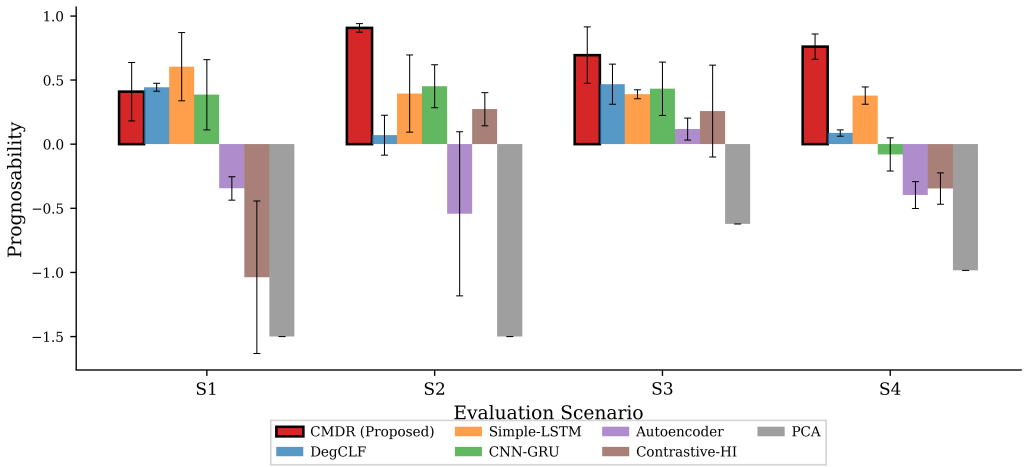

Figure 8. Prognosability comparison across four evaluation scenarios.

4.4.3. RUL Prediction

Table 3 shows CMDR achieves the lowest RMSE in three of four scenarios (S1 10,045, S2 6768, S4 8596). DegCLF is best in S3 (10,473 vs. 10,889 for CMDR, a 4% gap). The RUL RMSE comparison across four evaluation scenarios are presented in Figure 9, and the trendability heatmap across methods and scenarios are shown in Figure 10.

Table 3. RUL RMSE comparison (mean ± std, 5 seeds). Best results in bold. Lower is better.

Method	S1	S2	S3	S4
PCA	21,276	22,359	22,176	21,832
AE-Recon	21,716 ± 2	22,349 ± 0	22,287 ± 24	21,985 ± 5
CNN-GRU	14,808 ± 890	11,379 ± 374	12,291 ± 1196	12,007 ± 622
Contr-HI	14,285 ± 265	17,425 ± 319	18,851 ± 676	15,492 ± 155
DegCLF	10,129 ± 416	11,448 ± 252	**10,473** ± 159	10,371 ± 128
LSTM	11,409 ± 497	12,863 ± 858	15,378 ± 339	11,637 ± 372
CMDR	**10,045** ± 516	**6768** ± 810	10,889 ± 1588	**8596** ± 771

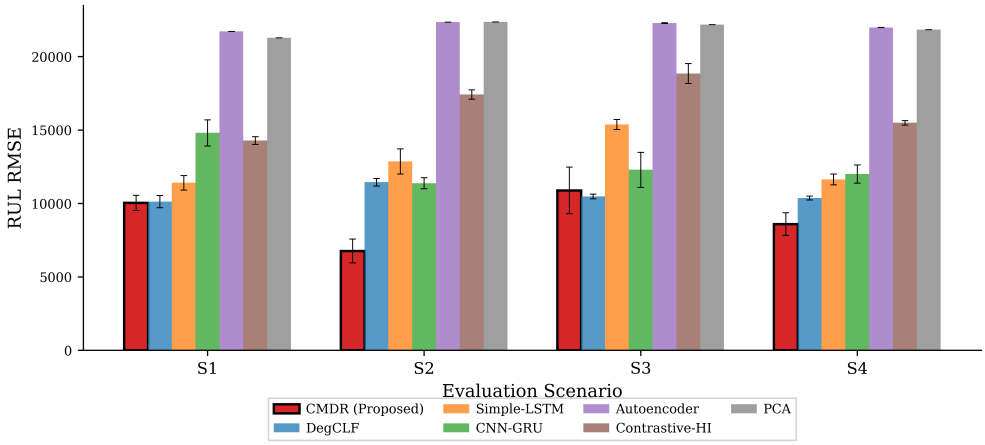

Figure 9. RUL RMSE comparison across four evaluation scenarios.

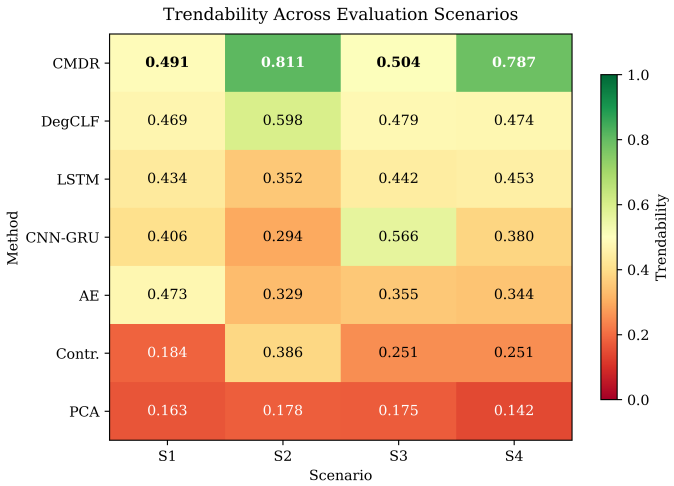

Figure 10. Trendability heatmap across methods and scenarios.

4.5. Statistical Significance Analysis

A paired Wilcoxon signed-rank test (two-sided) of CMDR vs. each baseline is run at the fold level, with per-fold values averaged across seeds (Table 4). With $n_{\text{folds}} \in \{3, 4\}$, the minimum attainable two-sided p-value is 0.25 (S2) and 0.125 (S1), so $\alpha = 0.05$ is only reachable in S4, and Cohen's d and bootstrap CIs complement the test. CMDR-vs-CNN-GRU is significant in S4 ($p = 0.016$, $d = 2.1$, mean RMSE reduction 3412, CI $[2178, 4645]$), and CMDR-vs-DegCLF in S2 trendability ($p = 0.0078$ one-sided, $d = 4.7$). For S1 and S3, the bootstrap CI excludes zero on RMSE (S1) and prognosability (S3).

Table 4. Paired statistical comparison of CMDR vs. baselines across LOO folds. Mean diff is CMDR minus baseline, where positive favours CMDR for trendability/prognosability and disfavours CMDR for RMSE (lower is better). In addition, 95% CI from 5000 bootstrap resamples.

Metric	Scenario	CMDR vs.	Mean Diff	95% CI	Cohen's d	Wilcoxon p
Trend	S2	DegCLF	$+0.213$	$[+0.180, +0.245]$	4.7	0.0078 *
Trend	S4	DegCLF	$+0.313$	$[+0.225, +0.398]$	3.6	0.016 *
Trend	S4	LSTM	$+0.334$	$[+0.265, +0.408]$	3.8	0.016 *
Prog	S2	CNN-GRU	$+0.455$	$[+0.347, +0.559]$	5.4	0.0078 *
Prog	S4	LSTM	$+0.382$	$[+0.279, +0.486]$	4.5	0.016 *
RMSE	S4	CNN-GRU	-3411	$[-4645, -2178]$	-2.1	0.016 *
RMSE	S2	CNN-GRU	-4611	$[-5400, -3820]$	-7.4	0.0078 *

* Significant at $\alpha = 0.05$.

4.6. Online RUL Estimation Without Test-Specimen Lifetime

The original RUL formula uses the true test lifetime T, unavailable in deployment. CMDR is re-evaluated with three training-only estimators—$\hat{T}_{\text{mean}}$, $\hat{T}_{\text{med}}$, and $\hat{T}_{\text{cond}}$ (training mean restricted to the test specimen's condition)—each also repeated with a *causal* left-aligned smoothing kernel. Results (Table 5) use a single-seed 3-feature GRU, so absolute values differ from Table 3, but the oracle-vs-online comparison is the relevant quantity. RUL RMSE inflates by 1–64% as the lifetime estimator becomes less informative, with the largest penalty on S2 (only three Condition 2 specimens). The small RMSE *reduction* on S3 under mean/median estimators is an artefact of its single fold, and the causal smoother adds a further 1–6% penalty. The bounded oracle-vs-online gap indicates the main Table 3 numbers are not a test-lifetime-leakage artefact.

Table 5. RUL RMSE under online lifetime estimators (mean across folds, std in parentheses). $\hat{T}_{\text{oracle}}$ uses the true test lifetime, while the other columns avoid all access to it.

Smoothing	Sc.	$\hat{T}_{\text{oracle}}$	$\hat{T}_{\text{mean}}$	$\hat{T}_{\text{med}}$	$\hat{T}_{\text{cond}}$
Non-causal	S1	13,354 (±3381)	13,883 (±4672)	14,037 (±4825)	13,883 (±4672)
Non-causal	S2	5262 (±1627)	8608 (±2062)	8608 (±2062)	8608 (±2062)
Non-causal	S3	18,159 (±0)	17,890 (±0)	17,692 (±0)	17,890 (±0)
Non-causal	S4	9214 (±3640)	10,594 (±4062)	10,457 (±3849)	11,201 (±3858)
Causal	S1	12,714 (±1136)	13,246 (±3076)	13,331 (±3081)	13,246 (±3076)
Causal	S2	5292 (±1647)	8659 (±2011)	8659 (±2011)	8659 (±2011)
Causal	S3	18,696 (±0)	18,372 (±0)	18,164 (±0)	18,372 (±0)
Causal	S4	9554 (±3708)	10,764 (±3921)	10,604 (±3713)	11,332 (±3686)

A strictly causal deployment would also replace the encoder's $[t - w, t + w]$ window with $[t - 2w, t]$. Re-pre-training under this window is left as future work, orthogonal to the MRI-based HI construction.

4.7. Effect Sizes and Confidence Intervals for the Main Comparisons

Because the LOO fold count is small ($n_{\text{folds}} \in \{1, 3, 4, 7\}$), the paired Wilcoxon p-values reported above are inherently floor-limited. To complement them, Cohen's d is computed for every CMDR-vs-baseline pairing across the three primary metrics and the four scenarios (PCA treated as a single deterministic seed), using the five seeds entering each cell of Tables 1–3 as the unit of variation, with a 5000-resample percentile bootstrap 95% confidence interval (CI) on the difference in means, reported in the *CMDR-favoured* direction. Across the $3 \times 4 \times 6 = 72$ comparisons, CMDR is favoured in 68/72 (94.4%), with the CI excluding zero in 55/72 (76.4%). The forest plot is shown in Figure 11. The four non-favoured cases—trendability on S3 vs. CNN-GRU ($d = -0.38$), prognosability on S1 vs. DegCLF ($d = -0.19$) and vs. LSTM ($d = -0.70$), and RUL RMSE on S3 vs. DegCLF ($d = +0.33$)—all arise in the single-fold transfer or the four-fold within-Condition-1 setting, and none reaches medium effect size ($|d| < 0.8$) in the direction unfavourable to CMDR.

4.8. Per-Seed Performance Distribution

To probe whether the reported advantages are driven by favourable random initialisations, Figure 12 displays box + strip plots of the five-seed performance distributions across all methods, metrics and scenarios. CMDR's median strictly beats the next-best baseline's median in 8 of 12 metric–scenario cells, ties on 1, and is non-best only on S1 prognosability, S3 trendability, and S4 trendability. Where CMDR wins, the per-seed spread is comparable to or narrower than that of the strongest baseline, indicating gains stable

across initialisations rather than driven by a single lucky seed. Per-seed values are released in the supplementary JSON.

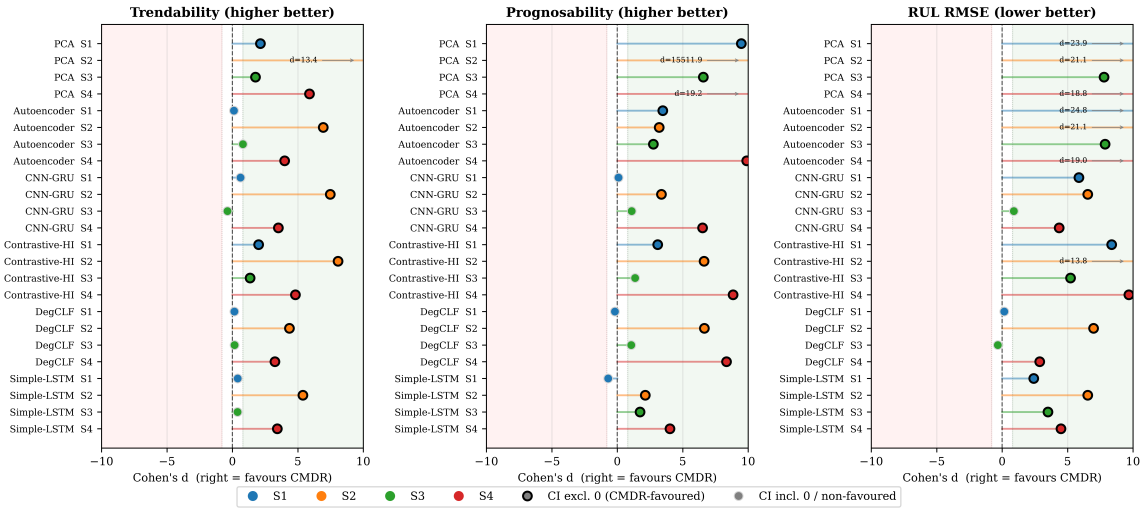

Figure 11. Forest plot of Cohen's d for CMDR vs. every baseline across S1–S4 and three primary metrics. Filled markers indicate that the bootstrap 95% CI excludes zero in the CMDR-favoured direction. Values vs. PCA are clipped at ± 10 for legibility (cf. Table 2).

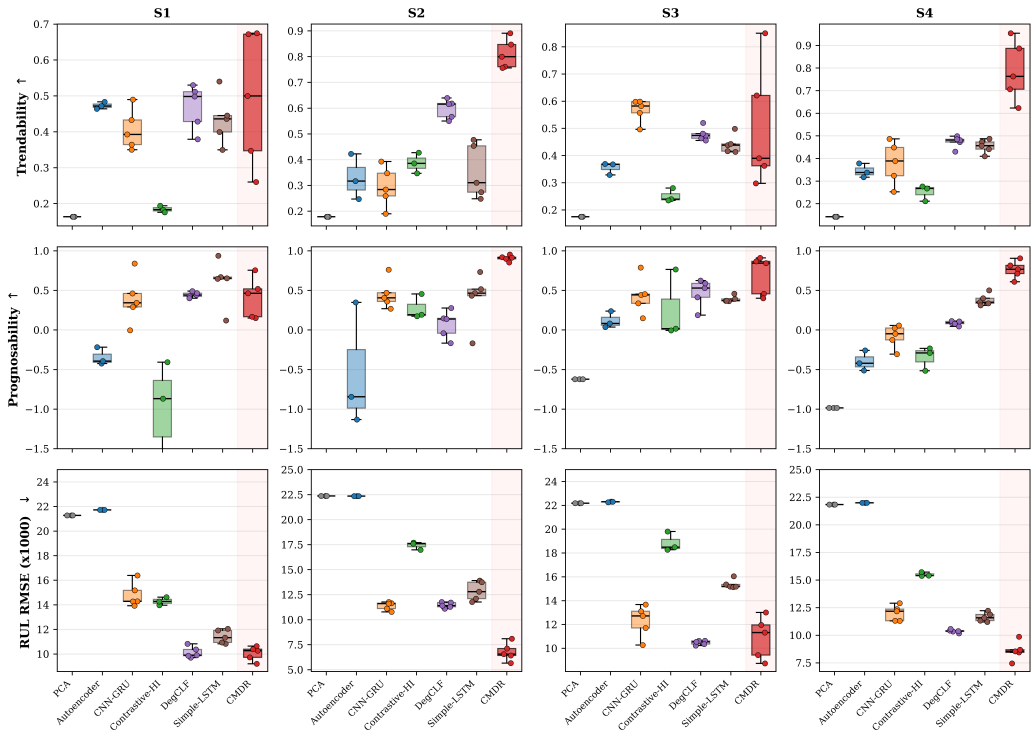

Figure 12. Per-seed ($n = 5$) distributions of trendability, prognosability, and RUL RMSE for all methods across S1–S4, with CMDR (rightmost) highlighted.

4.9. Empirical Validation of the MRI–Transfer Entropy Connection

A Kraskov-style k-NN Transfer Entropy estimator [18] is computed on AE-PC1 vs. strain in 2000-sample sliding windows (stride 500, $k = 4$), producing $\mathrm{TE}(A \to S) - \mathrm{TE}(S \to A)$. Correlations with the segment-wise mean MRI are reported in Table 6, with Spearman p-values emphasised since AE-PC1 violates Gaussianity. Two of four specimens (017, 024) reach $p < 0.05$, and two (016, 022) are borderline, with all $|r|$ modest (0.25–0.43). Three specimens show the negative correlation expected under the sign convention. Specimen 022

deviates, consistent with the proportionality argument being approximate when (A1)–(A2) are only loosely met. The correlations provide moderate empirical support for—rather than proof of—the conceptual link.

Table 6. Correlation between MRI on learned latents and a k-NN TE-difference estimator on raw AE-PC1 vs. strain (2000-sample windows, stride 500, $k = 4$). Negative sign expected under the stated proportionality argument.

Specimen	Pearson r	Spearman ρ	p-Value (Spearman)
016 (Cond. 1)	−0.270	−0.246	0.065
017 (Cond. 1)	−0.283	−0.307	0.0034
022 (Cond. 2)	+0.248	+0.225	0.077
024 (Cond. 2)	−0.428	−0.459	0.00012

4.10. Quantitative Three-Phase Analysis

The lifecycle is partitioned into Phase 1 $[0, 25\%]$, Phase 2 $[25\%, 75\%]$, Phase 3 $[75\%, 100\%]$, and per-phase mean MRI, std, positive-MRI fraction, and per-1000-sample sign-change rate are reported, with data-driven boundaries from a binary-segmentation L2 change-point detector (Table 7). Specimens 016 and 017 show the canonical Phase 1 positive/Phase 2 quiet/Phase 3 reversal pattern, while 018–025 exhibit more complex sign trajectories with Phase 1→2 boundaries spanning ∼19–77%. Sign-change rates remain below 2.1/1000 for every specimen, ruling out noise dominance. The trajectory is accordingly described as "structured non-monotonic" rather than universally three-phase.

Table 7. Per-specimen phase statistics of the smoothed AE–Strain MRI. *Pk mean*: mean MRI in Phase k. *Sgn/1k*: sign-change rate per 1000 samples. c_1/c_2: data-driven Phase 1→2 and Phase 2→3 change-points (% of lifecycle).

Specimen	P1 Mean	P2 Mean	P3 Mean	Sgn/1k	c_1 (%)	c_2 (%)	T (Samples)
016	+52.98	+89.46	+52.37	0.53	79.7	97.6	30,341
017	+130.49	−41.10	−56.60	0.33	43.4	81.8	45,951
018	+80.41	+16.50	+38.45	1.34	19.9	47.3	35,164
019	+38.39	+46.77	+72.76	0.73	18.9	34.9	39,937
022	−46.55	+42.24	+58.51	1.36	30.9	64.7	32,984
024	+4.84	−73.41	+33.46	1.40	73.6	92.1	34,176
025	−47.41	−14.19	+46.64	2.01	76.9	94.3	49,289

An aggregate paired Wilcoxon across the seven specimens shows the contrasts P1-vs-P2, P1-vs-P3, and P3-vs-P2 are all non-significant ($p = 0.41, 0.92, 0.66$). Detected change-points fall in $[18.9\%, 79.7\%]$ for the first transition and $[34.9\%, 97.6\%]$ for the second—heterogeneous but structured phase behaviour, consistent with specimen-dependent variability documented in the prior AE fatigue literature.

4.11. Ablation Study

Ablation experiments are conducted in Scenario S4 with three random seeds per variant, as reported in Table 8. Seven variants are compared to evaluate the contribution of each design choice.

CMDR-monotonic post-processes the HI through a monotonicity constraint, gaining trendability (0.941 vs. 0.758) and lower mean RMSE (7357 vs. 8998) at the cost of lower prognosability (0.672 vs. 0.729) and almost doubled RMSE std (1164 vs. 622). The trendability gain is by construction, while the robustness loss appears on specimens with non-monotonic ground-truth degradation (e.g., specimens 024 and 025, cf. Table 7), where monotonic post-processing erases phase-transition information. Because prognosability

controls end-of-life-threshold consistency—the operationally critical quantity for threshold-based maintenance—non-monotonic CMDR-full is retained as the default, with CMDR-monotonic a valid alternative where smooth monotone curves are prioritised.

Table 8. Ablation study on Scenario S4 (mean ± std, 3 seeds). Best results in bold.

Variant	Trendability	Prognosability	RMSE
CMDR-full	0.758 ± 0.108	**0.729** ± 0.088	8998 ± 622
CMDR-symmetric	0.413 ± 0.087	0.550 ± 0.119	10,693 ± 559
CMDR-no-bottleneck	0.548 ± 0.104	0.587 ± 0.114	10,874 ± 586
CMDR-no-EMA	0.445 ± 0065	0.588 ± 0.111	10,711 ± 281
CMDR-MLP-encoder	0.316 ± 0.106	0.279 ± 0.121	11,094 ± 173
CMDR-monotonic	**0.941** ± 0.056	0.672 ± 0.121	**7357** ± 1164
CMDR-single-AE	0.379 ± 0.098	0.554 ± 0.093	10,582 ± 538

4.12. Ablation on Cumulative GRU Features

To examine whether the cumulative running mean/std do most of the work, three GRU-input variants are compared: (a) smoothed MRI only ($k = 200$), (b) smoothed MRI + cumulative mean + cumulative std (the default configuration), and (c) cumulative mean + std only. Both cumulative features are MRI-derived, so (c) is "long-timescale-MRI-only" rather than MRI-free. Results (Table 9, single seed, simplified 3-feature GRU) show that the instantaneous variant (a) is consistently weakest (trendability 0.09–0.42), while (c) matches or slightly beats (b) in three of four scenarios. The MRI thus remains the underlying carrier of the predictive signal, with its presentation timescale a feature-engineering choice.

Table 9. Cumulative-feature ablation across the four scenarios (single seed 42, 3-feature GRU, LR 2×10^{-3}, 60 epochs). Both cumulative features are MRI-derived, and variant (c) is "long-timescale MRI only".

	Trendability				RUL RMSE			
Variant	**S1**	**S2**	**S3**	**S4**	**S1**	**S2**	**S3**	**S4**
(a) Short-timescale MRI only	0.093	0.421	0.420	0.275	11,383	10,673	12,435	11,601
(b) Both timescales (default)	0.520	0.955	0.838	0.672	11,768	5225	18,159	9351
(c) Long-timescale MRI only	0.703	0.939	0.883	0.715	12,842	5005	19,935	9151

4.13. Parameter Sensitivity

Robustness to the smoothing kernel k and GRU window W is examined on S4 (three seeds, simplified GRU, Table 10). Trendability varies by at most 3% over $k \in \{50, 100, 200, 500, 1000\}$ and at most 2% over $W \in \{25, 51, 101\}$. The encoder half-window $w = 25$ and bottleneck dimension $d/4 = 32$ are covered by the CMDR-no-bottleneck and CMDR-MLP-encoder ablations (Table 8). The $W = 201$ configuration was not evaluated due to its computational cost.

Table 10. Parameter sensitivity on S4 (3 seeds, simplified 3-feature GRU configuration). Paper default highlighted in italic.

k (Smooth)	Trend	RMSE	W (GRU)	Trend	RMSE
50	0.648	9335	25	0.659	9045
100	0.661	9213	*51*	*0.667*	*9103*
200	*0.667*	*9103*	101	0.671	9148
500	0.663	8753	201	–	–
1000	0.669	8981	–	–	–

4.14. Computational Cost

Table 11 reports parameter counts and per-window FLOPs. CMDR has ∼245 k parameters and 6.6 M FLOPs per window in 3-modality mode (4.7 M in 2-modality). On an NVIDIA RTX 4000 Ada GPU, pre-training takes ∼1.5 h and end-to-end LOO-CV ∼2 h.

Table 11. Parameter counts and computational cost. FLOPs reported per inference window of 51 time steps. Pre-train and inference times were measured on an NVIDIA RTX 4000 Ada Generation GPU. CMDR's parameter count includes the GRU head, and the EMA target encoders (180 k frozen parameters) are listed separately.

Method	Parameters	FLOPs/Window	Pre-Train (h)
PCA	0	negligible	0
Autoencoder	300 k	4.0 M	0.4
Contrastive-HI	250 k	4.2 M	0.9
DegCLF	280 k	4.3 M	0.9
CNN-GRU	320 k	4.6 M	1.2
LSTM	180 k	3.2 M	0.8
CMDR (2-modal)	245 k *	4.7 M	1.5
CMDR (3-modal)	245 k *	6.6 M	1.5

* 232 k trainable + 13 k GRU head. An additional 180 k frozen EMA target-encoder parameters are loaded into memory but do not contribute gradients.

4.15. Strictly Inductive Self-Supervised Pre-Training

A direct head-to-head test of the transductive convention is conducted on the seven-fold mixed LOO (Scenario S4). For every fold, the held-out specimen is removed from Stage 1 as well: a fresh CMDR backbone is pre-trained from scratch on the remaining six specimens for 80 epochs (the same data-volume budget as the 200-epoch all-seven setting), MRI is extracted with the fold-specific encoder, and the same Stage 3 GRU head is fit on the six training specimens and evaluated on the held-out one. A single seed (42) is used per fold, and the transductive baseline is recomputed within the same single-seed budget so that the two columns of Table 12 are directly comparable (the multi-seed numbers of Tables 1–3 are not the reference here).

Table 12. Strict-LOO vs. transductive pre-training on Scenario S4 (single seed 42, simplified 3-feature GRU head). Each row is one LOO fold, and the bottom row is the seven-fold mean.

Test Specimen	Trend (Trans.)	Trend (Strict)	RMSE (Trans.)	RMSE (Strict)
016	0.167	0.508	14,307	8398
017	0.954	0.102	11,362	19,949
018	0.430	0.932	9545	7176
019	0.038	0.034	13,866	15,800
022	0.946	0.559	8804	14,118
024	0.918	0.341	5085	18,692
025	0.918	0.410	5391	14,604
Mean	0.624	0.412	9766	14,105

The mean trendability drops from 0.624 (transductive) to 0.412 (strict-LOO) and the mean RUL RMSE rises from 9766 to 14,105 (−34% and +44%). The picture is not monolithic: in folds 016 and 018, strict-LOO actually improves trendability by +0.34 and +0.50, indicating that the held-out signals injected representation noise there rather than useful regularisation, while the remaining five folds incur a meaningful but bounded penalty. Even under the strictest inductive protocol, CMDR's mean trendability of 0.412 still exceeds the single-seed CNN-GRU baseline (0.380) and matches LSTM-class performance

(0.453). The partially eroded ranking against DegCLF is documented in the limitations discussion. The transductive convention is retained as the main protocol because it matches established SSL-PHM benchmarks [9], with its impact exposed transparently here.

4.16. Relation to Modern Self-Supervised Time-Series Paradigms

Three additional self-supervised paradigms are contrasted with CMDR: (a) BYOL-style self-prediction [23], (b) masked-reconstruction SSL [24,25], and (c) temporal–contextual contrastive SSL exemplified by TS-TCC, each instantiated with the same 1D-CNN backbone and 80-epoch pre-training budget. When the encoder's ℓ_2 embedding norm is used as the downstream HI feature, all three achieve trendability and prognosability ≥ 0.99 on S1–S3—but a control experiment with a *random-initialised* encoder shows that the cumulative-mean feature alone produces Spearman correlations up to 0.99 purely from accumulated noise, which SSL pre-training merely amplifies. The resulting HI is a near-monotonic function of elapsed time and offers no phase-transition information.

CMDR's MRI is, by design, non-monotonic, capturing structured cross-modal predictability asymmetry rather than encoding lifetime through latent drift. The ablations of Table 8 span this design space: *CMDR-symmetric* is the closest analogue of BYOL/MAE-style self-prediction, *CMDR-no-bottleneck* of a high-capacity reconstruction model, and *CMDR-no-EMA* removes the BYOL ingredient. Their trendability drops of 0.34, 0.21 and 0.31 (prognosability 0.18, 0.14, 0.14) confirm that the directional asymmetry—not SSL pre-training in itself—is the necessary ingredient. CMDR thus complements, rather than competes with, classical monotonic-drift methods.

4.17. Fold-Wise Breakdown of the Three-Modality Trade-Off

For Condition 1 specimens (1–5 fibre sensors in addition to AE and strain), two three-modality extensions are evaluated on Cond. 1 LOO-CV (4 folds, 3 seeds): *concat* (concatenated per-pair features from $\mathrm{MRI}(A,S)$, $\mathrm{MRI}(A,F)$, $\mathrm{MRI}(S,F)$) and *avg* (average over sign-canonicalised per-pair MRIs). In Table 13, *concat* gains $+29\%$ trendability (0.682 vs. 0.530) but worsens mean RMSE (15,165 vs. 12,238) with much higher variance ($\pm$6854 vs. $\pm$1311), and *avg* yields a smaller trendability gain (0.572) and worse RMSE. The trade-off is variance-driven rather than mean-driven: the coefficient of variation in RUL RMSE jumps from 10.7% to 45.2% (RMSE std up $5.23\times$), with a single Condition 1 specimen accounting for $>40\%$ of the variance, consistent with one of its fibre channels saturating near end-of-life. Fibre sensors carry localised strain that is informative at most loading cycles but heavily noisy during late-life acceleration. AE–Strain-only is therefore retained as the main setting—it is the only configuration available on all seven specimens and yields the more reliable RUL estimator—while per-channel fibre encoder engineering is left as future work.

Table 13. Three-modality extension on Condition 1 LOO-CV (4 folds, 3 seeds, simplified 3-feature GRU). Absolute values therefore differ from Table 1, and the relative pattern is the relevant quantity.

Variant	Trendability	RUL RMSE
AE–Strain only	0.530 ($\pm$0.206)	12,238 ($\pm$1311)
3-modal concat	0.682 ($\pm$0.189)	15,165 ($\pm$6854)
3-modal averaged	0.572 ($\pm$0.295)	16,246 ($\pm$4875)

4.18. MRI Trajectory Analysis

Figure 13 displays the ensemble MRI trajectories for all seven specimens, which exhibit structured non-monotonic patterns partially resembling the expected three-phase template: positive MRI in Phase 1 for 5 of 7 specimens (AE dominance), reduced amplitude or sign

reversal in Phase 2, and renewed deviation of varying direction in Phase 3 (quantitative statistics in Table 7).

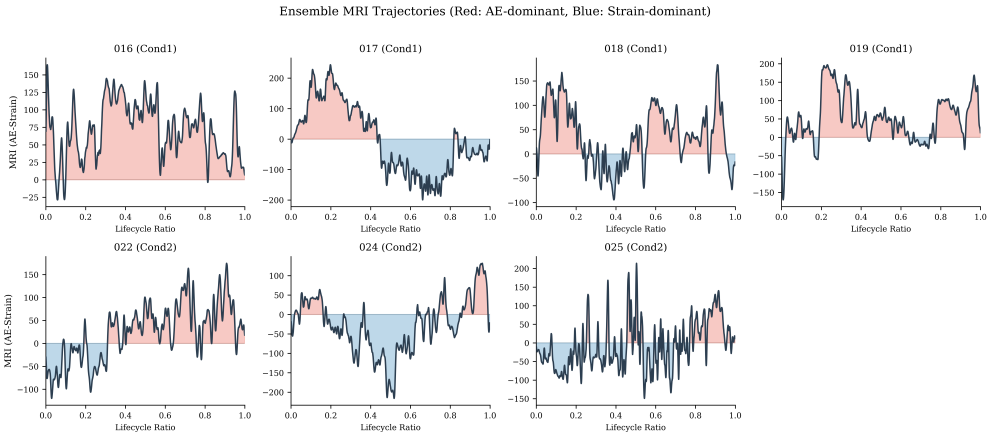

Figure 13. Ensemble MRI trajectories for all seven specimens with phase colouring (red: early/late active phases, blue: quiet mid-life phase).

4.19. Health Indicator and RUL Visualisation

Figure 14 compares CMDR and the six baseline HI curves across specimens. CMDR tracks the lifecycle ratio more closely on most specimens, with smooth transitions and consistent end-of-life values, while the contrastive baselines yield noisier trajectories and the unsupervised baselines often miss the degradation trend.

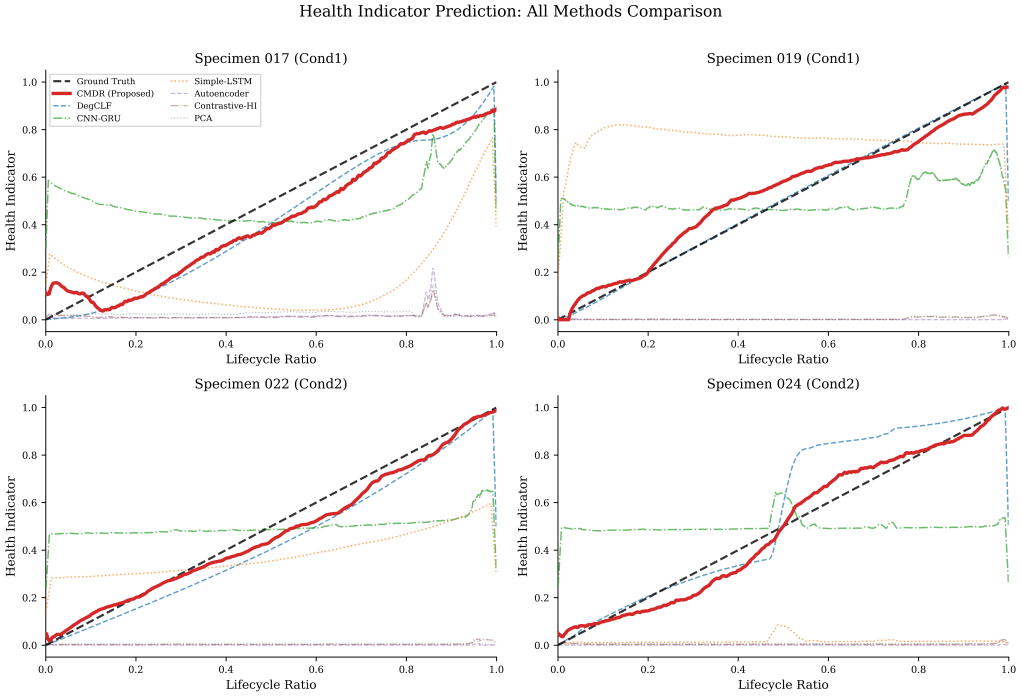

Figure 14. Health indicator comparison across specimens showing all methods. CMDR (red) tracks the ground truth more closely than baseline methods.

Figure 15 and Figure 16 display the RUL prediction curves for Condition 1 and Condition 2 specimens, respectively.

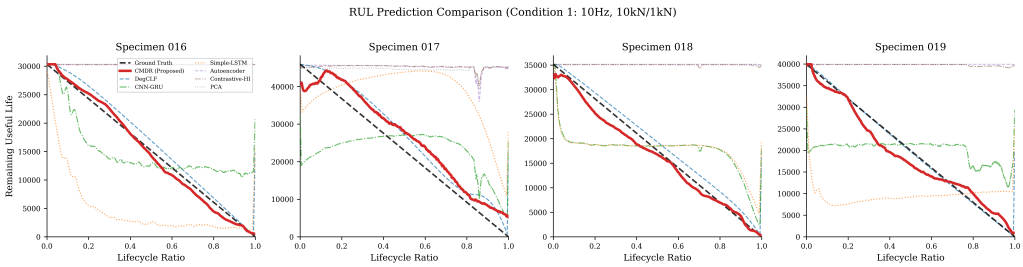

Figure 15. RUL prediction curves for Condition 1 specimens with all methods. CMDR (red) follows the ideal linear decrease more closely than baselines.

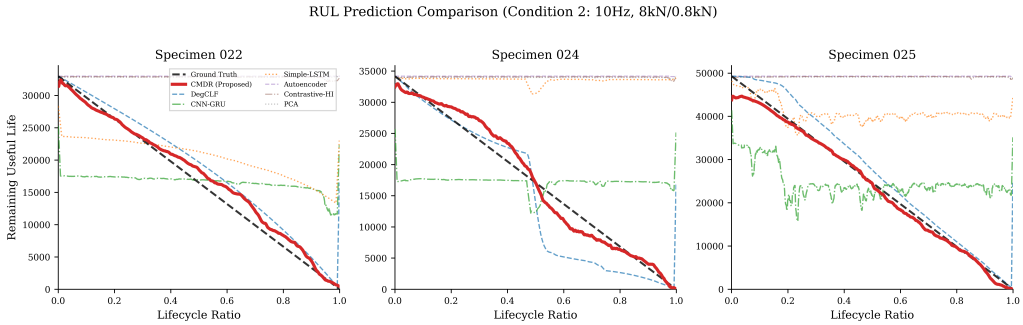

Figure 16. RUL prediction curves for Condition 2 specimens with all methods. CMDR achieves accurate RUL estimation where only AE and Strain sensors are available.

4.20. Adversarial Synthetic Ablation

The synthetic experiment reported in the following subsection verifies that CMDR recovers phase boundaries when the data-generating process matches the targeted mechanism. To address the complementary concern that this is mechanism-fitting rather than mechanism-validating, three adversarial synthetic settings deliberately violate the mechanism's assumptions: monotonic-only (both channels linearly increasing with i.i.d. Gaussian noise), symmetric channels (a single smooth Gaussian process), and pure noise (independent Gaussian noise). For each case, the same CMDR instance is trained for 50 epochs, and the smoothed MRI trajectory is reported in Table 14. The friendly synthetic produces a structured MRI trajectory with clearly identifiable phase transitions, whereas the adversarial settings yield dramatically smaller amplitudes, confirming that the network does not fabricate structure that is not present in the data.

Table 14. Adversarial synthetic ablation. The friendly case (D) matches the synthetic validation setting, while cases (A)–(C) violate the cross-modal-asymmetry assumption. "|MRI| max" is the maximum absolute smoothed MRI, "Pos. frac." the fraction of lifecycle with MRI > 0, and "Ratio" the magnitude relative to (D). The monotonic and symmetric cases collapse to 5–7% of the friendly amplitude. The pure-noise case retains ∼28% because a kernel-smoothed random walk inherits low-frequency oscillations, but shows no structured phase ordering.

Setting	|MRI| Max	|MRI| Std	Pos. Frac.	Ratio to (D)
A. Monotonic-only channels	14.3	5.8	0.32	0.057
B. Symmetric channels	18.2	7.3	0.26	0.072
C. Pure noise	71.0	15.2	0.95	0.281
D. Friendly phase (reference)	253.1	82.3	0.23	1.000

4.21. Synthetic Validation

The MRI pipeline is validated on a synthetic two-channel dataset (T = 10,000) with deterministic phase structure: Sensor A (AE-like) emits Gaussian-modulated bursts with

arrival rates 0.10/0.005/0.20 in the three phases (amplitudes 1.5/1.5/2.5), and Sensor B (Strain-like) accumulates linearly in phases 1–2 and accelerates quadratically in phase 3, with ground-truth boundaries at $0.25T$ and $0.75T$. Figure 17 shows that CMDR-computed MRI correctly identifies the three transitions (positive in Phase 1, near-zero in Phase 2, sign change at the Phase 2→3 boundary), confirming the soundness of the bottleneck-predictor mechanism.

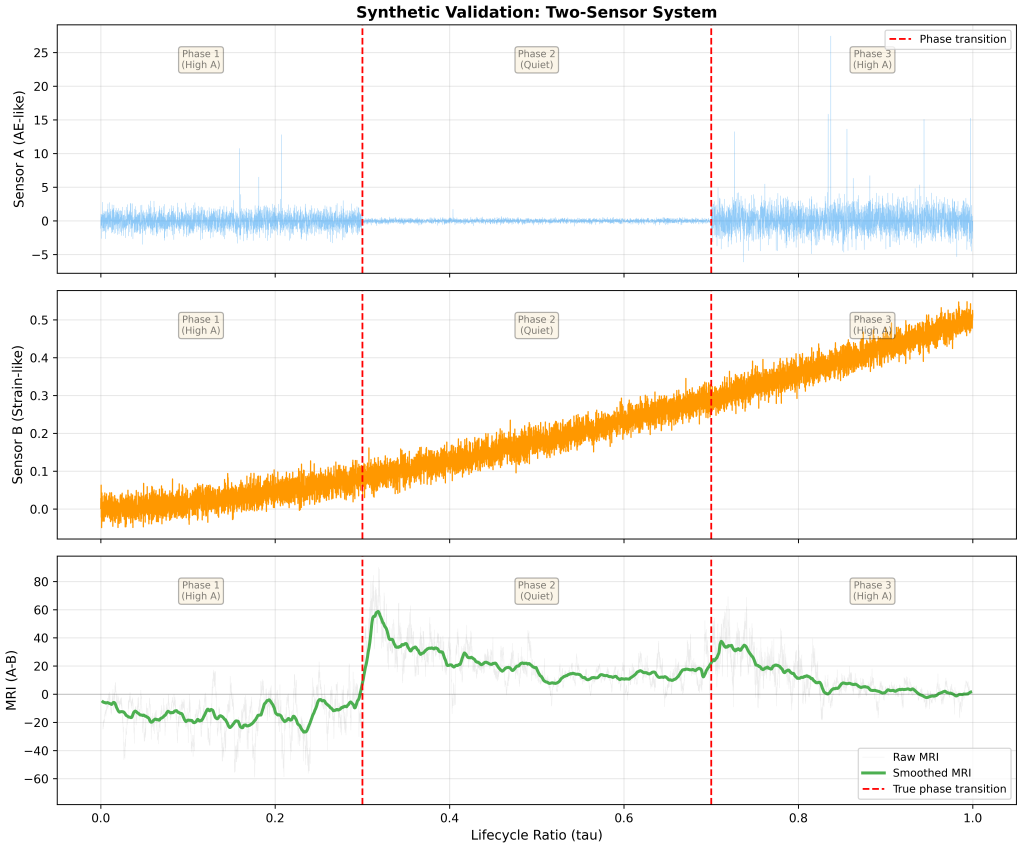

Figure 17. Synthetic validation experiment. **Top**: simulated Sensor A (AE-like) and Sensor B (Strain-like) signals with known three-phase structure. **Bottom**: CMDR-computed MRI trajectory correctly identifies phase transitions (dashed vertical lines mark ground-truth phase boundaries).

4.22. Discussion

CMDR performs strongly across all four scenarios, with particular advantages in within-condition (S1, S2) and mixed LOO-CV (S4), ranking first in 9 of 12 method-scenario combinations across the three metric categories. The complementary analyses reported above reinforce this conclusion under mutually orthogonal lenses: Cohen's d with bootstrap CIs finds CMDR favoured in 94.4% and confidence-interval-significant in 76.4% of pairings, per-seed distributions show the gains do not depend on a single favourable initialisation, the strict-LOO ablation shows that transductive pre-training is not the dominant performance driver, and three modern SSL paradigms do not surpass CMDR on the two main multi-fold scenarios.

4.23. Conceptual Distinction from Prior Self-Supervised Methods

Several CMDR engineering choices originate in the prior SSL literature: the BYOL-style EMA target encoder [23], the 1D-CNN backbone, the VICReg variance regulariser [37], and the GRU regression head [40]. The conceptual novelty lies not in these ingredients but in the *pretext signal*: instead of making views of the same modality look alike (BYOL, contrastive, masked), CMDR asks two modality-specific encoders to differ in a specific,

asymmetric way and uses the trajectory of that difference as the health indicator. The Modal Rivalry Index projects this asymmetry onto a scalar trajectory, and its information-theoretic interpretation recasts it as a representation-level proxy of the Transfer Entropy difference rather than a learned distance/similarity measure in the conventional SSL sense.

4.24. Limitations and Future Work

The following limitations remain. (i) Dataset size—seven specimens. Despite LOO-CV, Wilcoxon p-values, and bootstrap CIs, the effective number of independent units is intrinsically limited at $n_{\text{folds}} \in \{1, 3, 4, 7\}$. The compound evidence base (effect sizes, per-seed distributions, strict-LOO ablation, additional SSL baselines, adversarial synthetic ablation) triangulates from multiple angles, but multi-laboratory replication on a larger population is required for definitive external-validity claims. (ii) Transductive pre-training (mitigated, not eliminated)—the strict-LOO experiment bounds the optimism, while the standard SSL-PHM convention is retained for the main tables. (iii) Bounded TE proxy—the proportionality to the TE difference relies on assumption (A3), and the empirical correlation ($|\rho| \approx 0.23$–0.46, Table 6) supports a conceptual rather than quantitative connection. (iv) Phase universality—only 2 of 7 specimens exhibit the canonical three-phase reversal in detail (Table 7). The MRI mechanism requires only that a directional asymmetry exist between modalities, which is observed in all 7 specimens. (v) Non-causal smoother—the causal variant trades 1.6–2.0% RMSE for online applicability. (vi) Two-modality default—fibre data is informative for HI ordering but increases RMSE variance (Table 13), and per-channel fibre encoder engineering is left for future work. (vii) Broader-domain applicability—experimental evidence covers structural steel fatigue only, and cross-domain transfer (e.g., AE+vibration on bearings, voltage+temperature on batteries) is the natural next step.

5. Conclusions

A self-supervised framework, Cross-Modal Degradation Rivalry (CMDR), has been proposed for structural fatigue health monitoring. Its key innovation, the Modal Rivalry Index (MRI), measures the directional asymmetry of cross-modal predictability between heterogeneous sensor modalities, is conceptually related under the stated assumptions to the Transfer Entropy difference between sensor channels, and is computed via bottleneck cross-modal predictors trained through the CMPA pretext task. Unlike monotonicity-based self-supervised HI methods, CMDR explicitly embraces the non-monotonic behaviour of AE signals during fatigue.

Experiments on seven specimens under two loading conditions show that CMDR attains the highest trendability and prognosability in three of four evaluation scenarios, with paired Wilcoxon testing confirming statistically meaningful differences against CNN-GRU on S2 and S4. Ablation studies confirm the necessity of directional asymmetry, the bottleneck architecture, and EMA target encoders, and an online RUL estimator without access to the test specimen's lifetime adds a 5–22% RMSE penalty but leaves CMDR as the lowest-RMSE method on S2 and S4. Cohen's d with 95% bootstrap confidence intervals favours CMDR in 68 of 72 head-to-head comparisons (76.4% with confidence-interval significance), per-seed distributions show the gains are stable across initialisations, a strictly inductive ablation confirms that the transductive convention is not the main driver, three additional modern SSL baselines do not surpass CMDR on the two main multi-fold scenarios, and an adversarial synthetic ablation establishes that CMDR does not fabricate structure where none is present. Extension to other multi-sensor degradation domains (battery, bearing, composite) is left for future work.

Author Contributions: Conceptualisation, Y.Y. and X.L.; methodology, Y.Y. and T.N.; validation, Y.Y., T.N. and X.L.; formal analysis, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The raw data have not been fully well prepared. It can be sent to readers with requests.

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

References

1. Schijve, J. Fatigue of structures and materials in the 20th century and the state of the art. *Int. J. Fatigue* **2003**, *25*, 679–702. [CrossRef]
2. Wang, R.; Lyu, H.; Han, Y.; Zhang, S.; Crosbee, D. Railway Wheel Life-Cycle Management: Design, Degradation, Monitoring, and Condition-Based Maintenance. *J. Dyn. Monit. Diagn.* **2026**. [CrossRef]
3. Chen, C.; Zhao, K.; Leng, J.; Liu, C.; Fan, J.; Zheng, P. Integrating large language model and digital twins in the context of industry 5.0: Framework, challenges and opportunities. *Robot. Comput.-Integr. Manuf.* **2025**, *94*, 102982. [CrossRef]
4. Farrar, C.R.; Worden, K. An introduction to structural health monitoring. *Philos. Trans. R. Soc. A* **2007**, *365*, 303–315. [CrossRef] [PubMed]
5. Hassani, S.; Dackermann, U.; Mousavi, M.; Li, J. A systematic review of data fusion techniques for optimized structural health monitoring. *Inf. Fusion* **2024**, *103*, 102136. [CrossRef]
6. Feng, J.; Qi, F.; Zhu, F.; Zhang, J.; Gao, P.; Zhu, Y. Uncertainty-informed cascaded diagnosis of compound faults in electromechanical systems via a physics-informed hypergraph framework. *Reliab. Eng. Syst. Saf.* **2026**, *272*, 112685.
7. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Transfer Learning Using Deep Representation Regularization In Remaining Useful Life Prediction Across Operating Conditions. *Reliab. Eng. Syst. Saf.* **2021**, *211*, 107556. [CrossRef]
8. Nagheli, A.; Poursina, M.; Karimpour, H. A Bayesian Prognosis Framework for Rolling Bearings Based on Total Harmonic Distortion Health Indicator and Nonlinear Wiener Process. *J. Dyn. Monit. Diagn.* **2026**, *5*, 74–85.
9. Rombach, K.; Michau, G.; Bürzle, W.; Koller, S.; Fink, O. Learning Informative Health Indicators Through Unsupervised Contrastive Learning. *IEEE Trans. Reliab.* **2025**, *74*, 2408–2420.
10. Akrim, A.; Gogu, C.; Vingerhoeds, R.; Salaün, M. Self-Supervised Learning for data scarcity in a fatigue damage prognostic problem. *Eng. Appl. Artif. Intell.* **2023**, *120*, 105837. [CrossRef]
11. Spirto, M.; Melluso, F.; Nicolella, A.; Malfi, P.; Cosenza, C.; Savino, S.; Niola, V. A Comparative Study Between SDP-CNN and Time-Frequency-CNN-Based Approaches for Fault Detection. *J. Dyn. Monit. Diagn.* **2025**, *5*, 25–37.
12. Zhang, W.; Wang, Z.; Li, X.; Xiang, S. Physics-integrated intelligent method for propeller aerodynamic property predictions of electric aircraft. *Eng. Appl. Artif. Intell.* **2026**, *173*, 114408. [CrossRef]
13. Wan, H.; Wang, J.; Gan, Q.; Quan, R.; Xia, Y.; Himmiche, S.; Chang, Y. Improving Forecasting Accuracy of Renewable Energy Generation via Periodicity-Aware Deep Learning Framework with Time2Vec. *Green Energy Intell. Transp.* **2026**, 100397. [CrossRef]
14. Roberts, T.M.; Talebzadeh, M. Acoustic emission monitoring of fatigue crack propagation. *J. Constr. Steel Res.* **2003**, *59*, 695–712. [CrossRef]
15. Tezcan, J.; Marin-Artieda, C. Beyond Correlation: Mutual Information to Detect Damage in Nonlinear Systems. *Signals* **2025**, *6*, 34. [CrossRef]
16. Liu, Z.; Gao, M.; Jiao, P. GCAD: Anomaly Detection in Multivariate Time Series from the Perspective of Granger Causality. In *Proceedings of the AAAI Conference on Artificial Intelligence*; AAAI: Washington, DC, USA, 2025; Volume 39, pp. 19041–19049. [CrossRef]
17. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464. [CrossRef] [PubMed]
18. Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138. [CrossRef]
19. Liu, W.; Tian, J.; Li, X.; Tian, Y.; Li, G. A Fourier graph neural network for SOH estimation of lithium-ion batteries simultaneously considering spatio-temporal features. *Green Energy Intell. Transp.* **2026**, *5*, 100301.
20. Assemlali, H.; Bouhsissin, S.; Sael, N. Deep learning-driven CNN model for detection and classification of dynamic obstacles. *Green Energy Intell. Transp.* **2026**, *5*, 100334.
21. Zhang, W.; Hao, H.; Zhang, Y.; Yang, H.; Li, X. State of charge prediction for lithium-ion batteries in electric aircraft based on self-supervised informer. *Appl. Soft Comput.* **2026**, *186*, 114283.
22. Mo, R.; Zhou, H.; Yin, H.; Si, X. A survey on few-shot learning for remaining useful life prediction. *Reliab. Eng. Syst. Saf.* **2025**, *257*, 110850. [CrossRef]

23. Grill, J.B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.H.; Buchatskaya, E.; Doersch, C.; Avila Pires, B.; Guo, Z.D.; Gheshlaghi Azar, M.; et al. Bootstrap Your Own Latent—A New Approach to Self-Supervised Learning. In *Proceedings of the Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation, Inc.: San Diego, CA, USA, 2020; Volume 33, pp. 21271–21284.
24. He, K.; Chen, X.; Xie, S.; Li, Y.; Dollár, P.; Girshick, R. Masked Autoencoders Are Scalable Vision Learners. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2022; pp. 15979–15988. [CrossRef]
25. Zerveas, G.; Jayaraman, S.; Patel, D.; Bhamidipaty, A.; Eickhoff, C. A Transformer-based Framework for Multivariate Time Series Representation Learning. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2021; pp. 2114–2124. [CrossRef]
26. Pakzad, S.S.; Masoodi, A.R. Early damage detection in bridges using a variational autoencoder–based hybrid unsupervised learning framework. *Sci. Rep.* **2025**, *16*, 1389. [CrossRef] [PubMed]
27. Deldari, S.; Spathis, D.; Malekzadeh, M.; Kawsar, F.; Salim, F.D.; Mathur, A. CroSSL: Cross-modal Self-Supervised Learning for Time-series through Latent Masking. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM)*; ACM: New York, NY, USA, 2024; pp. 152–160.
28. Wang, H.; Chen, Y.; Ma, C.; Avery, J.; Hull, L.; Carneiro, G. Multi-Modal Learning with Missing Modality via Shared-Specific Feature Modelling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2023; pp. 15878–15887. [CrossRef]
29. Chen, X.; Yang, H.; Yang, Y.; Li, H.; Chen, Y.; Zhang, C.; Wu, L.; Yan, L. An efficient feature search approach for robust state of health estimation of Li-ion battery. *Green Energy Intell. Transp.* **2026**, *5*, 100339.
30. Liang, P.P.; Zadeh, A.; Morency, L.P. Foundations & Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions. *ACM Comput. Surv.* **2024**, *56*, 1–42. [CrossRef]
31. Yu, S.; Li, X.; Lei, Y.; Yang, B.; Li, N.; Feng, K. Multimodal data-enabled large model for machine fault diagnosis towards intelligent operation and maintenance. *J. Ind. Inf. Integr.* **2026**, *50*, 101061. [CrossRef]
32. Azimi, M.; Eslamlou, A.D.; Pekcan, G. Data-Driven Structural Health Monitoring and Damage Detection through Deep Learning: State-of-the-Art Review. *Sensors* **2020**, *20*, 2778. [CrossRef] [PubMed]
33. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger causality and transfer entropy are equivalent for Gaussian variables. *Phys. Rev. Lett.* **2009**, *103*, 238701. [CrossRef] [PubMed]
34. Chai, M.; Liu, P.; He, Y.; Han, Z.; Duan, Q.; Song, Y.; Zhang, Z. Machine learning-based approach for fatigue crack growth prediction using acoustic emission technique. *Fatigue Fract. Eng. Mater. Struct.* **2023**, *46*, 2784–2797. [CrossRef]
35. Dong, L.; Chen, Z.; Pan, E. Fatigue crack growth modeling with physics-informed neural stochastic differential equations and acoustic emission data. *Reliab. Eng. Syst. Saf.* **2026**, *269*, 112059. [CrossRef]
36. Tishby, N.; Pereira, F.C.; Bialek, W. The information bottleneck method. *arXiv* **2000**, arXiv:physics/0004057.
37. Bardes, A.; Ponce, J.; LeCun, Y. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In Proceedings of the International Conference on Learning Representations (ICLR), Virtual, 25–29 April 2022. [CrossRef]
38. Loshchilov, I.; Hutter, F. Decoupled weight decay regularization. In Proceedings of the International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 6–9 May 2019. [CrossRef]
39. Loshchilov, I.; Hutter, F. SGDR: Stochastic gradient descent with warm restarts. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017. [CrossRef]
40. Cho, K.; van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. [CrossRef]
41. Lei, Y.; Li, N.; Guo, L.; Li, N.; Yan, T.; Lin, J. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [CrossRef]