NDT
Journal of Non-Destructive Testing

Review

AI-Enabled Bridge Nondestructive Testing Using Vibration and Dynamic Response: An Investigation of Diagnostic Capability and Evidence Maturity

Muhammad Ziad Bacha, Mauro Sassu and Mario Lucio Puppio

Special Issue
AI-Powered Nondestructive Testing and Condition Assessment for Infrastructure

Edited by
Dr. Amine El Mahdi Safhi, Dr. Mahdi Kioumarsi and Dr. Afaq Ahmad

MDPI

https://doi.org/10.3390/ndt4040033

NDT
Journal of Non-Destructive Testing

MDPI

Review

AI-Enabled Bridge Nondestructive Testing Using Vibration and Dynamic Response: An Investigation of Diagnostic Capability and Evidence Maturity

Muhammad Ziad Bacha *, Mauro Sassu and Mario Lucio Puppio

Department of Civil, Environmental Engineering and Architecture (DICAAR), University of Cagliari, Via Marengo 2, 09123 Cagliari, Italy; msassu@unica.it (M.S.); mariol.puppio@unica.it (M.L.P.)
* Correspondence: muhammadziad.bacha@unica.it

Abstract

Artificial intelligence (AI) is used to interpret vibration and dynamic-response measurements for bridge nondestructive testing (NDT) and structural health monitoring (SHM), but accuracy on a dataset does not show that a method is ready for use on real bridges. Searches of Scopus, Web of Science, and IEEE Xplore in August 2026 and a recall audit yielded 794 journal studies, each coded for diagnostic capability, validation realism, damage authenticity, generalisation, temporal maturity, engineering-use maturity, and five trustworthiness attributes. Enabling inference and detection account for 67.8% of studies and prognosis for 0.6%. Numerical validation is the largest category (45.5%); in-service validation covers 30.7%. Confirmed full-scale physical change appears in 61 studies, but only six (0.8%) combine it with in-service monitoring. Transfer to an unseen bridge appears in 16 studies (2.0%), year-scale evidence in 20 (2.5%), and maintenance decision support in one (0.1%); 65.2% show none of the trustworthiness attributes. These patterns held across ten sensitivity scenarios, including full-text verification of 52 studies. Agreement between the original screening and an independent human screening of 350 records was moderate ($\kappa = 0.43$), so some eligible studies may be missing. Diagnostic capability is broad, but the evidence needed for deployment remains scarce.

Keywords: bridge nondestructive testing; structural health monitoring; artificial intelligence; dynamic response; damage diagnosis; domain adaptation; physics-informed machine learning; uncertainty quantification

Check for updates

Academic Editors: Amine El Mahdi Safhi, Mahdi Kioumarsi and Afaq Ahmad

Received: 1 September 2026
Revised: 10 October 2026
Accepted: 21 October 2026
Published: 2 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.

1. Introduction

Bridges are long-life assets exposed to ageing, repeated traffic loading, thermal and environmental actions, fatigue, corrosion, scour, and support changes [1–4]. Visual inspection and conventional nondestructive testing (NDT) are performed at discrete intervals and cannot follow changes in structural behaviour between campaigns. Structural health monitoring (SHM) fills this gap by recording structural response under operational or controlled excitation and flagging changes that require engineering attention.

In this review, AI-enabled bridge NDT means nondestructive inference of structural condition from vibration and dynamic-response measurements, including continuous SHM [1–3]. Adjacent fields such as AI-assisted ground-penetrating radar (GPR) [5], computer vision, ultrasonic and impact-echo testing, thermography, and acoustic emission lie outside the systematic core. Multi-sensor vibration diagnosis of bridge components, such as the jack-arch study in [6], falls within it.

Vibration-based assessment is indirect. Acceleration, dynamic strain, displacement, modal parameters, and vehicle–bridge interaction (VBI) responses depend on mass, stiffness, damping, boundary conditions, excitation, and sensor characteristics [3,4]. Temperature, traffic, wind, road or track condition, and sensor drift can change them as much as early deterioration does [7–9]. An AI model must therefore separate damage-sensitive information from environmental and operational variability (EOV), and showing that it does so on a real bridge is much harder than showing that it classifies a dataset.

Machine learning (ML) on engineered features has been extended to deep learning on raw or transformed responses, healthy-only anomaly detection, self-supervised and open-set learning, transfer learning and domain adaptation, physics-informed learning, uncertainty quantification (UQ), explainable AI (XAI), information fusion, and digital twins [1–4]. Existing reviews classify this literature by algorithm, sensing modality, or application [1–4], as did the authors' earlier scoping review of dynamic-response-based monitoring [10]. Di Mucci et al. [1], for example, map AI methods across inspection, monitoring, and deterioration prediction and note limited validation and generalisation in qualitative terms. None of these reviews measure how strongly structural evidence supports each claimed capability. A model can reach near-perfect accuracy on numerically generated damage and still be far from operational use, whereas a less accurate model tested on a full-scale bridge with documented physical change provides stronger engineering evidence [11]. The same distinction applies to field data, transfer learning, response forecasting, and digital twins [12–14].

This review measures that evidence. Its contribution is methodological: no new algorithm is proposed, and no accuracy metrics are pooled. Every study in a complete systematic corpus is coded on six ordinal evidence dimensions (diagnostic capability, validation realism, damage authenticity, generalisation, temporal maturity, and engineering-use maturity) and five independent trustworthiness attributes, and the stability, accuracy, and sensitivity of the coding are tested. To the authors' knowledge, evidence maturity has not been measured in this way for AI-enabled bridge assessment. Six research questions (RQs) are addressed. RQ1 is answered descriptively in Section 3 and RQ2–RQ6 from the coded evidence in Sections 4–8; Figure 1 shows which evidence dimensions answer each question:

RQ1. *What vibration and dynamic-response information is used by AI-enabled bridge NDT, and how is that information represented for learning?*

RQ2. *What diagnostic capabilities, from enabling inference to detection, localisation, characterisation, and prognosis, have been demonstrated?*

RQ3. *How does demonstrated diagnostic capability compare with validation realism and physical damage authenticity?*

RQ4. *How effectively do existing approaches address environmental and operational variability, domain shift, and cross-bridge generalisation?*

RQ5. *To what extent are physics, uncertainty quantification, explainability, and information fusion incorporated into trustworthy AI-enabled bridge assessment?*

RQ6. *How far have current systems progressed from algorithmic outputs toward inspection support, deterioration prognosis, maintenance planning, and asset-management decisions?*

The review finds that diagnostic capability is broad, but that studies combining in-service validation, confirmed damage, transfer to an unseen bridge, long-term data, quantified trustworthiness, and a defined maintenance decision remain exceptional.

Table 1. Evidence framework: coding levels and boundary rules for diagnostic capability, validation realism, damage authenticity, generalisation, target supervision, temporal maturity, engineering-use maturity, and the five trustworthiness attributes.

Dimension	Code	Definition	Does not Qualify/Boundary Rule
Diagnostic capability	D0–D4	D0 enabling inference only; D1 detection; D2 localisation; D3 characterisation or quantitative diagnosis; D4 prognosis of future condition, damage, or remaining life	Prediction of future response is D0 unless future structural condition is estimated; response forecasting is not prognosis
Validation realism	R1–R4	R1 numerical/synthetic; R2 laboratory/reduced-scale; R3 full-scale bridge or benchmark in a dedicated controlled campaign; R4 in-service monitoring under operating conditions	Dedicated full-scale or benchmark campaigns (e.g., Z24, Old ADA, S101) are R3, not R4
Damage authenticity	A0–A4	A0 no confirmed physical damage; A1 numerical/synthetic damage; A2 physical laboratory damage; A3 synthetic/injected damage on real bridge measurements; A4 confirmed full-scale physical change	In-service data alone do not imply A4; real measurements combined with injected or simulated damage are A3
Generalisation	G0–G3	G0 same structure/domain; G1 unseen conditions on the same bridge; G2 transfer to, or validation on, an unseen bridge; G3 transfer across a population of distinct bridges	Transfer-learning terminology, simulation-to-experiment transfer on one structure, generic pretraining, extensive retraining on the target, or repeated application without a held-out source-to-target test do not qualify for G2/G3
Target supervision	ZS, UT, FS, ST	Zero-shot; unsupervised target information; few-shot labelled target data; substantial supervised retraining	—
Temporal maturity	T0–T2	Evaluated data span: T0 < 1 month; T1 $\geq$ 1 month and < 1 year; T2 $\geq$ 1 year (retrospective or prospective noted in Table S1)	Coded by the span of data actually evaluated; unreported durations are T0
Engineering-use maturity	M0–M3	M0 algorithmic output; M1 diagnostic information; M2 explicit warning, inspection, or intervention support; M3 prognosis linked to a defined maintenance or asset-management decision	M3 requires a defined decision (intervention threshold, inspection schedule, or load limit); a generic statement that results could support maintenance is neither M2 nor M3
Environmental and operational robustness	E	EOV modelled, compensated, normalised, or used to challenge the inference	Generic noise tests
Uncertainty quantification	U	Posterior or predictive distributions, intervals, uncertainty propagation, or failure probability of the inferred quantity	RMSE or Bayesian hyperparameter tuning alone
Explainability or physical interpretation	X	Demonstrated explainability or physically interpretable reasoning	An attention layer alone
Physics integration	P	Structural physics inside the inference	Physics used only to generate training data
Information fusion	F	Explicit, evaluated integration of separate sensor, variable, domain, model, or decision streams	Concatenation or a serial hybrid network

Source: authors' evidence-maturity framework (Section 2.4). Trustworthiness attributes are recorded as present or absent and are not combined into a score.

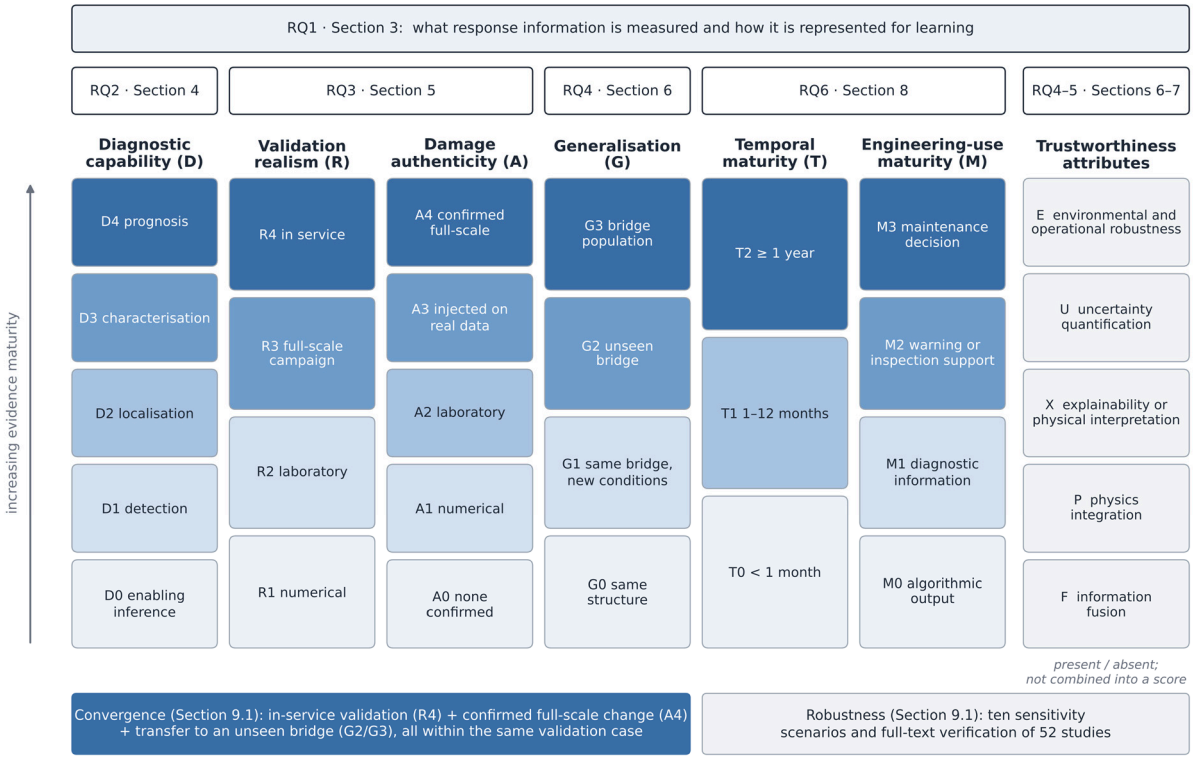


Figure 1. Structure of the review. Each research question (RQ) is answered from the evidence dimensions beneath it, and the section that reports it is shown above. Levels run from low (bottom) to high (top) evidence maturity; the five trustworthiness attributes are recorded as present or absent and are not combined into a score. Full definitions and boundary rules are given in Table 1. Progressively darker blue shading represents increasing evidence maturity.

2. Materials and Methods

2.1. Review Design

The study is a structured critical review based on a systematic literature search and multidimensional evidence coding, reported following the applicable PRISMA 2020 items [15]. The protocol was not registered; the eligibility criteria and codebook are given in Section 2.2 and Table 1, and every coding change is logged in Supplementary Table S3. Predictive performance was not meta-analysed. The studies differ in structural typology, sensing, excitation, damage definition, learning task, validation design, and metric, so pooled accuracy would imply a comparability that does not exist. The evidence dimensions in Section 2.4 serve as the appraisal of included studies: they grade how far each study's evidence supports its claimed capability, which risk-of-bias tools developed for clinical trials are not designed to do for engineering ML studies.

The first author defined the criteria and the codebook and supervised a large language model (ChatGPT 4.8, OpenAI, San Francisco, CA, USA) that assisted screening, eligibility assessment, and coding. The first author checked the workflow and resolved the boundary cases it flagged. Decisions and codes were recorded in spreadsheet ledgers, and duplicates were identified by DOI and normalised title. Three checks were added to test this workflow: a recall audit of excluded records, an independent human screening of a 350-record sample, and a full-text verification of 52 studies (Section 2.3).

2.2. Search Strategy and Eligibility

The systematic core comprises English-language peer-reviewed journal articles published from 1 January 2015 to the eligibility cutoff of 23 August 2026 (Supplementary Table S7). Scopus was searched on 23 August 2026, and Web of Science Core Collection and IEEE Xplore

on 24 August 2026. A publisher-designated citable version with a permanent DOI had to be available by the cutoff; repository-only preprints were excluded.

The query combined four concept blocks (bridge structure × dynamic response × AI × structural assessment); the full strings are given in Supplementary Table S5. Backward and forward citation tracing and targeted searches supplemented the queries for transfer learning, self-supervised and open-set learning, physics-informed methods, UQ, XAI, fusion, digital twins, and prognosis. Digital-twin studies were retrieved through a separate query that required bridge, dynamic-response, and AI or diagnostic content, so that architecture papers without AI-based structural inference were not imported.

Core primary studies had to (i) concern a bridge, a bridge component, or a structural model developed for bridge monitoring or damage assessment; (ii) use a dynamic-response quantity (acceleration, dynamic strain, displacement, modal information, or vehicle-induced response); (iii) include an AI or ML component; (iv) address structural inference, anomaly or damage detection, condition classification, localisation, prognosis, or a directly enabling SHM function (response reconstruction, virtual sensing, modal identification, or load estimation); (v) report enough validation information for coding; and (vi) be peer-reviewed journal articles. Studies whose application domain was generic structures, with a bridge as one of several test cases, were excluded (bridge-primary boundary). Visual-only defect detection, satellite interferometric (InSAR) displacement studies, sensor-fault classification without condition inference, inspection-database-only deterioration models, and seismic studies without monitoring data were also excluded. Reviews were cited as context only.

The searches returned 3412 records. After removal of 1745 duplicates and addition of 12 records from targeted searches, 1679 records were screened by title and abstract, 839 were assessed at full text, and 727 met the core criteria. The recall audit (Section 2.3) added 67 studies, giving a final core of 794 studies on which all prevalence statistics are based (Figure 2).

2.3. Coding and Quality Checks

Coding. Evidence was extracted on the structural system, sensing, response variable, representation, learning paradigm, diagnostic target, validation setting, damage provenance, EOV treatment, transfer, monitoring duration, uncertainty, explainability, physics integration, fusion, and engineering output. Each validation case within a study was examined separately. The unit of reporting is the publication (one record per study in Supplementary Table S1), coded at the highest level demonstrated for each dimension. Combinations such as R4+A4 are counted only when every criterion is met within the same validation case. A paper with an in-service healthy-monitoring case (R4, A0) and a separate benchmark damage case (R3, A4) is therefore counted under R4 and under A4, but not under R4+A4.

Coding stability. After first-pass coding of the original 727 studies, 336 studies were reconciled: 232 flagged for source checking and 104 selected for consistency or high-maturity audits. Codes changed in 277 of these 336 studies. A subsequent corpus-wide audit of the trustworthiness attributes changed further codes, so that in total 398 studies differ between first-pass and final codes. First-pass versus final agreement (quadratic-weighted Cohen's κ for ordinal dimensions, unweighted κ for attributes) was D 0.82, R 0.63, A 0.88, G 0.80, T 0.98, M 0.80, E 0.54, U 0.33, X 1.00, P 0.92, and F 0.45 (Supplementary Table S2). These values measure the stability of one workflow, not agreement between independent coders. The low values for E, U, and F reflect the strict final rules, which removed codes based on noise tests (E 268 → 148), probabilistic terminology

or prediction error (U 144 → 38), and hybrid architectures without demonstrated fusion (F 88 → 28) (Section 9.2; Supplementary Table S3).

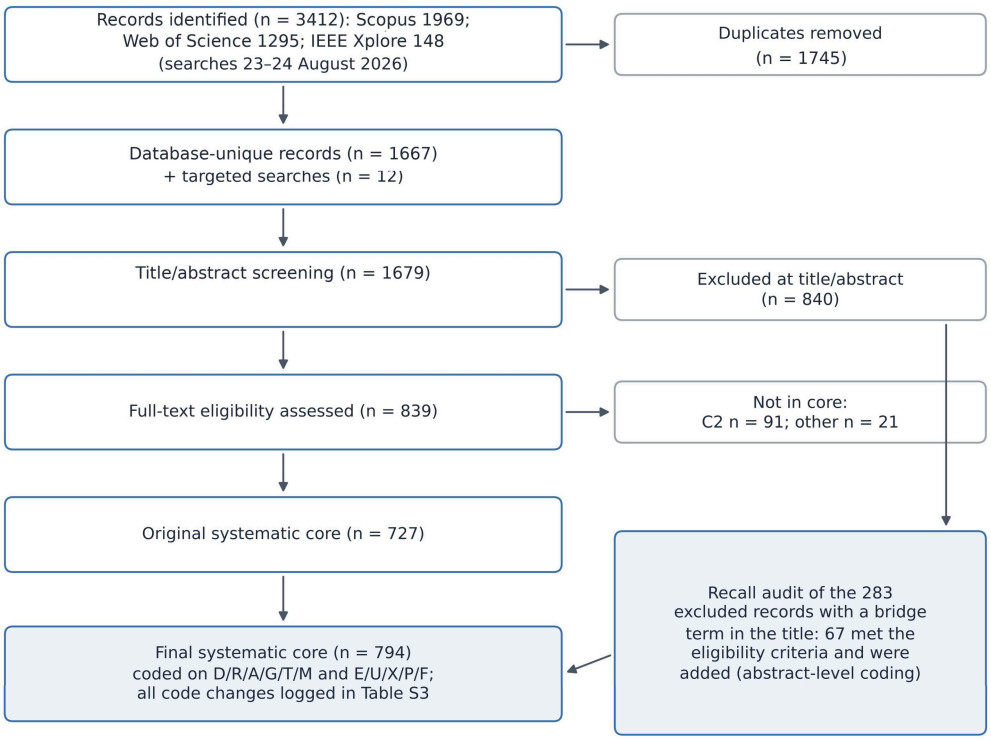


Figure 2. Corpus construction (searches 23–24 August 2026; eligibility cutoff 23 August 2026). Of 3412 records, 1679 were screened and 839 assessed at full text, giving 727 core studies; a recall audit of 283 bridge-titled exclusions added 67, giving a final core of 794.

Recall audit. A check of the title/abstract exclusions showed that a study of a bridge component [6] had been excluded under the rule 'primary application domain is not a bridge'. All 283 excluded records with a bridge term in the title (34% of the 840 exclusions) were therefore re-screened with the assistance of a second language model (Claude Opus 5.5, Anthropic, San Francisco, CA, USA), and 67 of them (24%) met the criteria and were added. Most additions are enabling (D0) studies of response prediction, reconstruction, or load identification on in-service bridges; they were coded from their abstracts and are flagged in Table S1. The 300 exclusions that mention a bridge only in the abstract were not re-screened, because the bridge-primary boundary excludes them by design.

Independent screening check. To test the title/abstract decisions, the first author screened a random sample of 350 records (21% of those screened) in Rayyan, without access to the original decisions and without having seen the records before. Two calibration rounds on a separate set of 250 records preceded this check and were used to clarify the wording of the criteria (Supplementary Table S6). Agreement with the original decisions was $\kappa = 0.43$ (95% CI 0.33–0.53; 72% raw agreement). The human screener was more inclusive: 89 of the 98 disagreements were records that the original screening had excluded (37 of them marked 'maybe'), whereas 164 of the 166 sampled records in the final core were also included by the human screener.

Full-text verification. To estimate coding accuracy, 52 studies were re-coded from their full texts with Claude Opus 5.5, without access to the existing codes. The sample had two parts: 8 studies from a stratified random draw of 50, of which the other 42 were not retrievable in full text; and 44 studies drawn at random from the 142 corpus studies in gold open-access journals, with high-maturity codes oversampled. Each of the 105 code-level disagreements with Table S1 (in 43 studies) was re-adjudicated from the

full text, and each verified code is documented with its source and a supporting quotation (Supplementary Table S6). Agreement with Table S1 was κ = 0.72 (D), 0.61 (R), 0.63 (A), 0.77 (G), 0.52 (T), 0.10 (M), 0.64 (E), 0.17 (U), 0.54 (X), −0.05 (P), and 0.65 (F), with raw agreement of 69–96%. The low κ for U and P reflects rare positive codes (at most 7 of 52), not frequent disagreement (87–88% raw agreement). The disagreements were systematic. Table S1 shows under-coded severity estimation (D1/D2 to D3 in 10 studies), evaluated damage (A0 to A1–A4 in 9), and multi-month spans (T0 to T1/T2 in 7), and over-coded decision support (none of the 7 M2 codes was confirmed). The verified codes are listed alongside the original codes in Table S1 and are tested as a sensitivity scenario (Section 9.1).

2.4. Evidence Dimensions and Trustworthiness Attributes

Table 1 defines the coding levels and lists, for each, what does not qualify. The following boundary rules decide the cases that recur. Prediction of future response is D0 unless future structural condition is estimated. In-service data alone do not imply A4, and real measurements combined with injected or simulated damage are A3. R3 denotes dedicated full-scale or benchmark campaigns such as Z24, Old ADA, or S101. Neither transfer-learning terminology nor transfer from a numerical model to an experimental version of the same structure qualifies for G2. Temporal maturity is coded by the span of the data actually evaluated, and unreported durations are coded T0. A single retrospective evaluation on several years of archived data therefore receives the same T2 code as repeated evaluation during multi-year monitoring; Table S1 records which of the two applies. M3 requires prognosis linked to a defined decision, such as an intervention threshold, inspection schedule, or load limit; a generic statement that results could support maintenance is neither M2 nor M3.

The five trustworthiness attributes are recorded as present or absent and are not combined into a score. E requires EOV to be modelled, compensated, normalised, or used to challenge the inference; generic noise tests do not qualify. U requires posterior or predictive distributions, intervals, uncertainty propagation, or failure probability; root-mean-square error (RMSE) or Bayesian hyperparameter tuning alone does not qualify. X requires demonstrated explainability or physically interpretable reasoning; an attention layer alone does not qualify. P requires structural physics inside the inference, not only in the generation of training data. F requires an explicit, evaluated integration of separate sensor, variable, domain, model, or decision streams; concatenation or a serial hybrid network does not qualify. Supplementary Tables S1, S3 and S4 give the study-level codes, boundary decisions, and sensitivity analyses.

3. Structural-Dynamic Basis and AI Paradigms

For a bridge represented in linearised form, the measured response follows:

$$M\,\ddot{u}(t) + C\,\dot{u}(t) + K\,u(t) = f(t), \quad (1)$$

where M, C, and K are the mass, damping, and stiffness matrices and f(t) is the excitation. Damage changes stiffness, load transfer, boundary conditions, damping, or mass. Traffic, temperature, wind, road or track condition, support behaviour, sensor state, and excitation change the response too. AI-based assessment is therefore an inverse structural-dynamics problem: condition must be inferred from response without assuming that every response change is damage.

3.1. Response Information and Representation

What the sensing system makes observable bounds what any AI model can infer. Acceleration is the most common input [11,16,17]; dynamic strain carries local information

on deformation and fatigue demand [18–20]; displacement captures low-frequency and quasi-static behaviour [21–24]; and modal quantities condense measurements into system-level descriptors [25–27]. In drive-by sensing, bridge information is embedded in the coupled vehicle–bridge response [7,16,28–31]. Malekjafarian et al. formulated the problem with numerical VBI data [16], whereas Duran et al. classified several states of a full-scale bridge subjected to controlled physical changes from vehicle-mounted acceleration [11].

Response data reach learning algorithms in three forms: engineered features such as modal parameters, statistical descriptors, and damage indices [27,32,33]; frequency or time-frequency representations [34–36]; and raw response histories [11,37,38]. Here, 'AI-ready information' means response data that have been acquired, preprocessed, aligned, and represented for ML inference; the term does not imply that the data suffice for diagnosis. The representation affects transfer as well as within-domain performance. Raw responses preserve subtle information but also encode sensor and excitation specifics [39,40]; modal features are interpretable but vary with geometry, boundary conditions, and environment [8,41]; and time-frequency features depend on windowing and normalisation [34,36].

Automated modal identification, virtual sensing [42,43], missing-data reconstruction [44–48], and response forecasting [49–51] are enabling functions (D0) unless a validated step converts their output into structural condition [25,27,52]. The attention-based convolutional neural network (CNN) of Haghbin et al., for example, predicts modal properties and is examined with XAI methods, but its output remains modal identification [53]. Environmental and operational variables provide the context needed to judge whether a change is structurally meaningful: AI systems can normalise features, condition baselines on the operating state, or flag observations outside the development distribution [7–9,24,26,54,55].

Fusion terms are used in a restricted sense. Multi-sensor fusion combines different sensors [6,56]; multi-physical-variable fusion combines quantities such as acceleration and strain [57,58]; multi-domain fusion combines alternative representations of the same measurement [13,35]; and feature- or decision-level fusion combines learned representations or model outputs [56,59]. Combining raw acceleration with time-, frequency-, and wavelet-domain features [13] is therefore multi-domain fusion, not multimodal sensing.

For RQ1, then, acceleration is the main input, with dynamic strain, displacement, modal quantities, and vehicle-borne responses as the others; these reach the model as engineered features, frequency or time-frequency images, or raw histories. The choice of representation governs how far a model can transfer, as well as how accurate it is on its own data.

3.2. Learning Paradigms

Supervised learning maps responses to predefined structural states, using classifiers on engineered features [60–62] or deep networks on raw or transformed signals [37,38,63]. Labelled damage states are rarely available from in-service bridges, so most supervised studies rely on numerical simulation, laboratory structures, or controlled full-scale campaigns [64–66]. Controlled full-scale studies show that supervised classification can go beyond numerical proof of concept [11,66,67], but the reported accuracy remains conditional on the structure, excitation, damage sequence, and sensor layout.

Healthy-only and unsupervised methods model normal behaviour and flag deviations [68–72], building on earlier novelty-based bridge detection [73–75]. Park and Kim trained a convolutional autoencoder with deep support vector data description on intact data and evaluated it on a damaged in situ steel truss bridge [76]; Giglioni et al. evaluated autoencoder novelty against progressive full-scale benchmark changes [71]. An

anomaly can also arise from environmental change, sensor malfunction, maintenance, or an unseen operating state [77–79], so these methods support detection (D1); localisation, mechanism, and severity require further evidence. Self-supervised learning builds representations from unlabelled data [80], and open-set learning removes the assumption that every future observation belongs to a trained class [81]. Open-set classification has been evaluated on Z24 data [81], and physics-encoded unsupervised learning on numerical and laboratory bridge networks [82].

Transfer learning and domain adaptation address the fact that a model developed on one structure rarely meets the same statistical and physical domain on another [41]. Ferreira et al. adapted healthy-state information from one bridge to a nearly identical neighbouring bridge without a healthy baseline, using Bayesian finite-element calibration [12]. Transfer is classified here by the target information used: zero-shot [36,83], unsupervised adaptation [12,40,84–86], few-shot fine-tuning [87], or supervised retraining. Transfer from a numerical model to an experimental version of the same structure [88] is not cross-bridge transfer. Population-based SHM extends transfer to groups of related assets, using laboratory populations [89] and numerical bridge networks [90].

Physics-informed ML constrains or augments data-driven inference with structural knowledge through governing equations, finite-element models, grey-box relationships, architectural constraints, or physics-informed loss functions [2,42,91]. Examples include physics-encoded unsupervised learning for bridge networks [82], physics-informed stiffness identification and early warning on an in-service railway bridge [18], and grey-box thermal modelling of a monitored viaduct [24].

Across these paradigms the data assumptions relax step by step, from fully labelled bridge-specific classification to healthy-only, self-supervised, open-set, transfer, population-level, and physics-guided inference [12,82,85,90]. A newer paradigm adds evidential value only if its capability holds under independent and physically representative validation, and Sections 4–8 measure how often that has been shown.

4. Results I: Diagnostic Capability

Figure 3 summarises the distributions for all six dimensions and the number of trustworthiness attributes per study; this section covers diagnostic capability. Enabling functions (D0) account for 280 studies (35.3%): modal identification [25,53,92], virtual sensing [42,43], response reconstruction and imputation [44–47,93,94], load and cable-force estimation [95–99], damping and energy-dissipation estimation [100,101], and response forecasting [49–51]. Detection (D1) accounts for 258 studies (32.5%) [71,76,77] and localisation (D2) for 103 (13.0%) [21,32,58]. Estimation of damage type, severity, or condition parameters (D3) accounts for 148 studies (18.6%) [37,60,64], of which 75 (50.7%) are validated only numerically (Figure 4).

Quantitative diagnosis on an operating bridge exists but is rare. Khan et al. reconstructed a spatial stiffness field on an in-service railway bridge with an inspection-confirmed fatigue crack, localised the affected region, quantified the stiffness reduction under measured temperature variation, and issued a warning 15 days before the existing monitoring system [18]. In terms of Table 1, this is quantitative diagnosis with in-service validation and confirmed damage (D3/R4/A4), evaluated over less than a month (T0), with warning support (M2) and three trustworthiness attributes (E, X, P).

Prognosis of future condition (D4) is demonstrated in five studies (0.6%): fatigue-damage prognosis of orthotropic steel decks from field-measured strain [20], physics-informed fatigue-crack evolution [102], fatigue-life prediction of bridge cables [103] and of coastal bridge details [104], and corrosion-fatigue life estimation of an in-service steel

bridge [105]. Response forecasting, however long its horizon, is not prognosis, because it predicts the response rather than the structural condition.

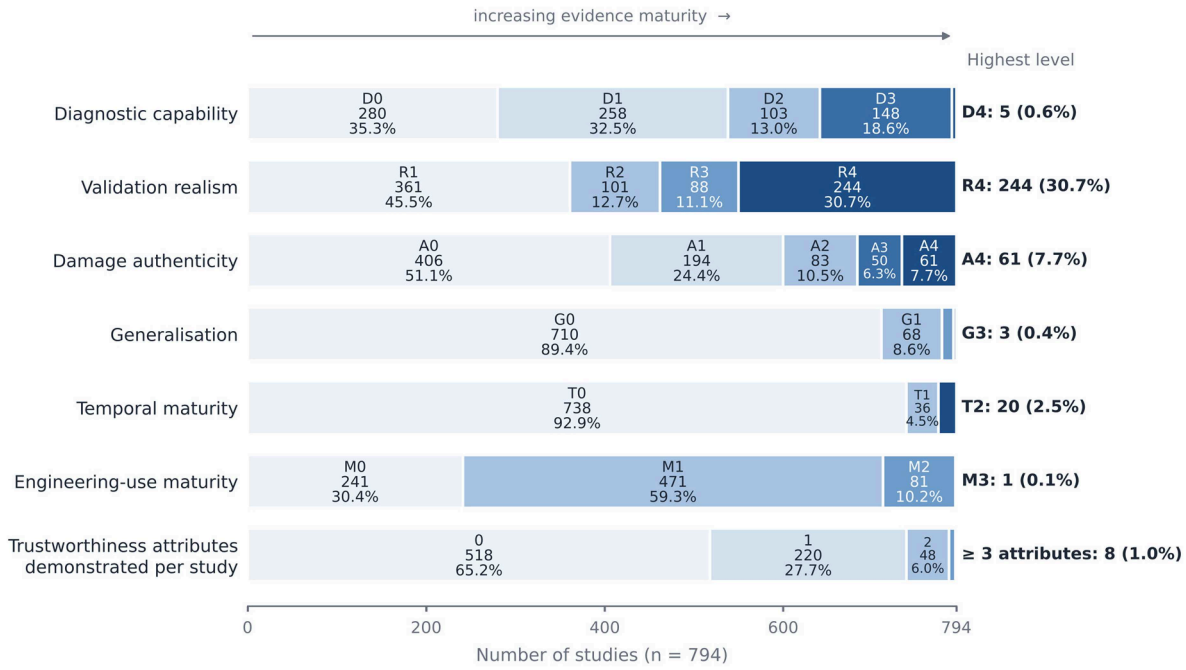


Figure 3. Evidence-maturity profile of the 794 systematic-core studies. Each bar shows how the studies are distributed across the levels of one dimension, from lowest (left) to highest (right) maturity; the bottom bar shows the number of trustworthiness attributes demonstrated per study. Segment labels give the number and percentage of studies (segments too narrow to label: D4, 5; G2, 13, G3, 3; T2, 20; M3, 1; 3 attributes, 7; 4 attributes, 1; 5 attributes, 0). The right-hand column gives the highest level and its share of the corpus. Source: Supplementary Table S1. Blue shading progresses from lighter (lower evidence levels) to darker (higher evidence levels).

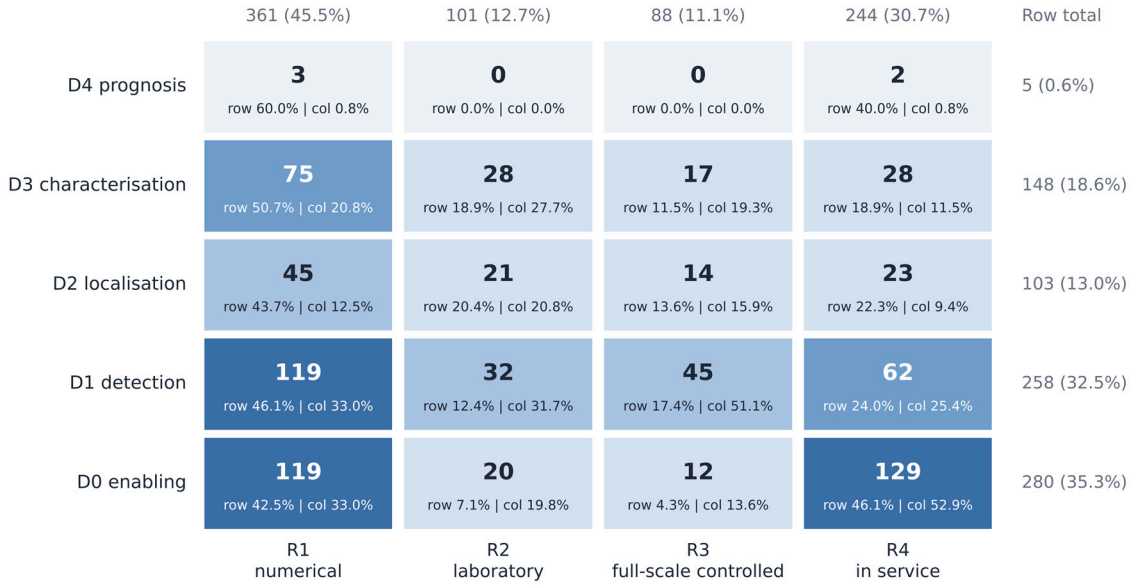


Figure 4. Diagnostic capability versus validation realism (n = 794). Each cell gives the number of studies and its percentage of the row (capability level) and of the column (realism level); margins give totals and percentages of the corpus. Cell shading increases with the number of studies.

Diagnostic capability (RQ2) is therefore broad but concentrated in enabling inference and detection (D0 and D1: 538 studies, 67.8%), and prognosis is exceptional. The D scale

orders only the structural question answered, not the strength of the evidence: a long-term, uncertainty-aware detector (D1) on an in-service bridge can provide stronger deployment evidence than a numerically validated prognosis model (D4).

5. Results II: Validation Realism and Damage Authenticity

Validation realism and damage authenticity are coded independently of capability, because a performance metric is meaningful only relative to the problem on which it was obtained. High accuracy on numerical damage scenarios can coexist with failure under real traffic, temperature, sensor drift, and structural uncertainty [106].

Numerical or synthetic validation (R1) is the largest realism category (361 studies, 45.5%), followed by in-service monitoring (R4; 244, 30.7%), laboratory or reduced-scale tests (R2; 101, 12.7%), and dedicated full-scale or benchmark campaigns (R3; 88, 11.1%), such as the progressive damage tests on Z24, Old ADA, and S101. R4 describes the realism of the setting, not of the damage. Of the 244 R4 studies, 181 involve no confirmed physical damage, 50 combine field measurements with injected or simulated damage (A3), 5 use numerical and 2 laboratory damage, and 6 involve confirmed full-scale change (A4).

Damage authenticity is distributed as A0 406 (51.1%), A1 194 (24.4%), A2 83 (10.5%), A3 50 (6.3%), and A4 61 (7.7%). The A3 category prevents a common misreading. Operational strain, acceleration, or modal data combined with finite-element (FE) damage, injected anomalies, or transferred synthetic labels [24,107,108] are useful for method development, but they do not constitute validation against authenticated damage.

Confirmed full-scale change (A4) is concentrated on a few structures. Of the 61 A4 studies, 55 are controlled campaigns (R3), 52 use a named benchmark bridge, and 40 use Z24 alone; counting each named bridge once leaves 13 publications (Supplementary Table S4). Six studies combine A4 with in-service monitoring in the same validation case (Table 2): a fatigue crack on a steel railway bridge [18], a damaged twin bridge [12], a cracking event on a bascule railway bridge [77], existing cracks monitored for four years on the Caohekou Bridge [109], field-measured scour on an in-service bridge [110], and damage localisation on a railway continuous-beam bridge [21].

Table 2. The six studies that combine in-service validation (R4) with confirmed full-scale physical change (A4) in the same validation case.

Ref.	Structure and Confirmed Change	D	G (Target)	T (Span)	M	Attributes
[18]	Steel railway bridge; inspection-confirmed fatigue crack	D3	G0	T0 (19 days)	M2	E, X, P
[12]	Twin concrete bridges; newer bridge reported as damaged, no healthy baseline	D1	G2 (UT)	T0	M1	none
[77]	Bascule railway bridge; cracking event	D1	G0	T0	M1	none
[109]	Caohekou Bridge; existing cracks	D3	G0	T2 (4 years)	M2	E
[110]	In-service bridge; field-measured scour	D3	G0	T0	M2	E
[21]	Railway continuous-beam bridge; damage localised from displacement influence lines	D2	G0	T0	M2	E, P, F

Source: Supplementary Table S1 (all six coded A4-confirmed). Codes as defined in Table 1; UT, unsupervised target information. Spans in parentheses are those reported; T0 without a span means the duration was not reported or was under one month.

Figure 4 cross-tabulates diagnostic capability against validation realism. D3 occurs under R1, R2, R3, and R4 in 75, 28, 17, and 28 studies, and D4 comprises three R1 and two R4 studies. Advanced capability and realistic validation seldom coincide (RQ3). What limits

the evidence is whether advanced outputs hold under realistic validation with authentic structural change.

6. Results III: Environmental Robustness and Generalisation

Temperature alters stiffness, support behaviour, and modal properties [8,26]; traffic changes excitation amplitude and frequency content; and wind, moisture, boundary changes, train characteristics, road roughness, and sensor replacement shift the measurement distribution further [7,17,34]. These effects can exceed the change caused by early deterioration [9,111]. The corpus includes in-service physics-aware thermal compensation [24], long-term strain modelling with environmental inputs [19,112], and drive-by detection tested under several EOV factors [17] and under road roughness with temperature variation [7]. Nevertheless, 710 studies (89.4%) remain within the development structure and domain (G0), and 68 (8.6%) test unseen conditions on the same structure (G1) (Figure 3).

Transfer to, or independent validation on, a bridge not used for development (G2) is demonstrated in 13 studies (1.6%), and transfer across a population of distinct bridges (G3) in three (0.4%) [90,113,114]. Four kinds of transfer were not coded as G2 or G3: simulation-to-experiment adaptation on one structure, as for the reduced-scale cable-stayed bridge of Lu et al. [88]; generic pretraining; extensive retraining on the target; and repeated application to several bridges without a held-out source-to-target test. Of the 16 G2/G3 studies, four are zero-shot, eight use unlabelled target data, three use few-shot labels, and one does not report enough to classify; none requires supervised retraining. Five are validated in service, five in controlled full-scale campaigns, and six numerically or in the laboratory, so five studies (0.6%) combine G2/G3 with in-service validation.

Transfer is a structural problem as well as a statistical one. Two datasets can be mathematically alignable while representing structures with different modal density, span arrangement, boundary conditions, sensors, or deterioration mechanisms [41,90]. Yano et al. frame the applicability of transfer learning for bridge SHM in terms of structural-domain compatibility [41], and the twin-bridge study shows how structural similarity and model uncertainty enter a practical source–target problem [12].

RQ4 has a two-part answer. EOV is addressed in a minority of studies (Section 7), and generalisation beyond the development bridge is rarer still: nine in ten studies stay on that bridge, and verified transfer to another bridge in service is reported in five.

7. Results IV: Trustworthiness

Environmental and operational robustness (E) is demonstrated in 173 studies (21.8%) [7–9], uncertainty quantification (U) in 44 (5.5%) [93,95,100,115–117], explainability or physical interpretation (X) in 43 (5.4%) [18,53], physics integration (P) in 48 (6.0%) [18,24,82,102], and information fusion (F) in 33 (4.2%) [6,13,56,59]. Of the 794 studies, 518 (65.2%) demonstrate none of the five attributes, 220 (27.7%) demonstrate one, 48 (6.0%) demonstrate two, seven (0.9%) demonstrate three, and one (0.1%) demonstrates four; none demonstrate all five (Figure 5).

The most frequent pairs are E+U (17 studies), X+P (12), and E+X (11), whereas U and X occur together in only two studies. The only four-attribute study is a probabilistic cable-tension monitoring framework on an in-service cable-stayed bridge (E, U, P, F) [95]. Three-attribute studies include the in-service railway bridge with a confirmed fatigue crack (E, X, P) [18], grey-box early warning on a monitored viaduct (E, X, P) [24], baseline-free damage localisation on a railway bridge (E, P, F) [21], and physics-guided fusion for scour monitoring (E, P, F) [118].

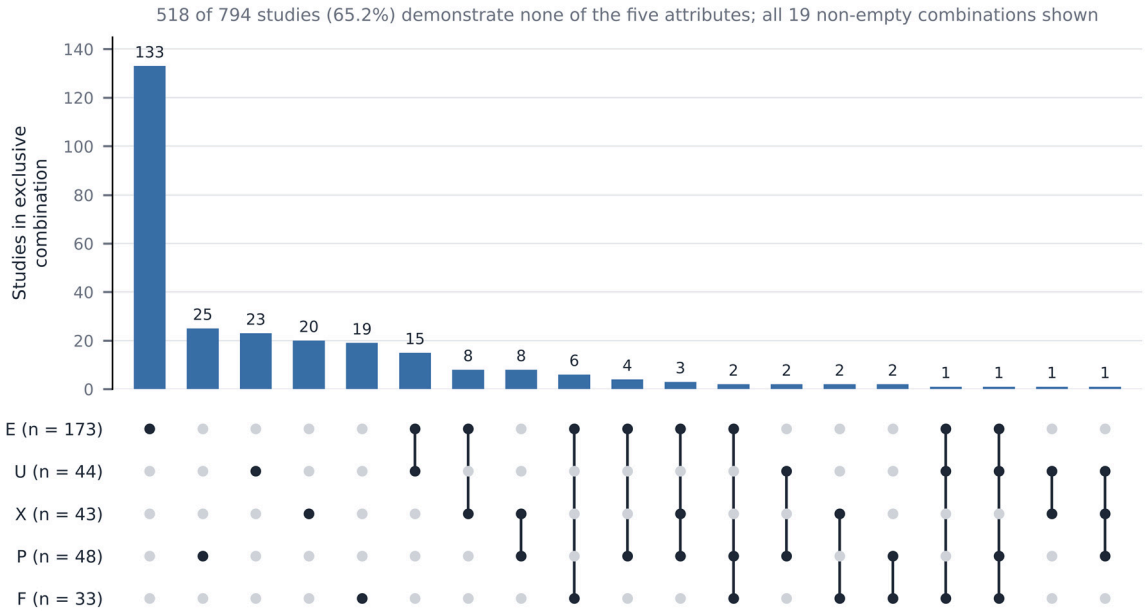


Figure 5. Exclusive co-occurrence of the five trustworthiness attributes (n = 794). E, environmental/operational robustness; U, uncertainty quantification; X, explainability/physical interpretation; P, physics integration; F, information fusion. Bars give the number of studies in each of the 19 non-empty exclusive combinations; row labels give the total for each attribute. A missing attribute means that qualifying evidence was not found under the coding rules. Filled dots indicate attributes present in a particular combination; pale dots indicate absent attributes. Connecting lines link attributes belonging to the same combination.

Trustworthiness, the subject of RQ5, is far less developed than predictive performance, and studies that combine several attributes under realistic validation are exceptional.

8. Results V: Temporal and Engineering-Use Maturity

Year-scale temporal evidence (T2) is reported in 20 studies (2.5%) and month-scale evidence (T1) in 36 (4.5%); the remaining 738 (92.9%) evaluate less than a month of data (T0). Of the 20 T2 studies, 13 evaluated archived monitoring data retrospectively and none reported prospective operation of a fixed model on newly acquired data; for seven, the accessible text did not allow classification (Supplementary Table S1).

Digital-twin terminology is not a maturity category. Bridge digital twins couple measurements, computational models, and analytical algorithms to maintain an evolving model of the asset [14,91,119,120]. In-service twins support model calibration and updating without condition prognosis [14,120–123], whereas AI-enabled twins also perform diagnosis [13,65]. A digital twin is therefore coded by its demonstrated output, from D0/M0 (response prediction and model updating) to D4/M3 (prognosis linked to a maintenance decision).

Early warning and prognosis are also distinct. Early-warning systems identify a developing abnormal condition earlier than an existing threshold or inspection [24,124,125]. The physics-informed system of Khan et al. warns before the existing system based on a stiffness change associated with a confirmed fatigue crack, but it does not forecast the crack state and is therefore M2, not D4 [18]. Response forecasting and reconstruction [49–51,126] remain D0. Xie and Bai demonstrate mechanism-informed prognosis by representing crack evolution and displacement degradation of orthotropic deck details in a physics-informed network, with numerical validation only [102]. The digital twin of the Espartxo Bridge links monitored behaviour with calibrated numerical scour scenarios for early scour detection [127]; its scour states are simulated, not observed.

The engineering-use distribution is M0, 241 (30.4%); M1, 471 (59.3%); M2, 81 (10.2%); and M3, 1 (0.1%). In the full-text verification sample (Section 2.3), none of the seven M2 codes were confirmed: the studies produced enabling or diagnostic output, but no warning, inspection, or intervention step was built and evaluated. The M2 share is therefore an upper bound.

One study (0.1%) reaches M3: a corrosion-fatigue life analysis used to set a truck axle-weight limit for a target reliability [105]. The other four D4 studies forecast condition without a defined decision layer and are not M3; the bridge-cable fatigue-life prediction [103], for example, compares the predicted life with the service life but defines no maintenance action. A deck-condition framework [59], originally coded D4/M3, was recoded D3/M1 after full-text checking: it predicts current condition ratings and forecasts strain and stress responses, not future structural condition, and its maintenance actions are suggested from rating bands without being evaluated. A monitoring study on the Caohekou Bridge [109], originally coded D4/M3, was recoded D3/M2: it predicts crack width and responses from environmental and operational conditions and supports early warning, but it neither forecasts future deterioration nor links it to a defined maintenance action.

On RQ6, few systems have progressed beyond algorithmic or diagnostic output: 20 studies reach T2, five reach D4, and one reaches M3, and none of the classifiable T2 studies operated a fixed model prospectively.

9. Discussion

The same pattern runs through all six dimensions. Numerical validation is the largest realism category, and prognosis, cross-bridge transfer, year-scale evaluation, authenticated in-service damage, and maintenance decision support each occur in less than 3% of studies (Figure 3). Section 9.1 tests whether this pattern is robust, Section 9.2 examines how the trustworthiness attributes interact, and Sections 9.3 and 9.4 set out what the results mean for bridge owners and where the review is limited.

9.1. Evidence Convergence and Sensitivity

Convergence. Of the six studies that combine in-service monitoring with confirmed full-scale change (Section 5), one, the twin-bridge study of Ferreira et al. [12], also demonstrates transfer to an unseen bridge (Figure 6). The newer of two nearly identical neighbouring bridges had been reported as damaged for almost a decade and had no healthy baseline. The authors calibrated an FE model of its undamaged state by Bayesian inference and used unsupervised domain adaptation to reuse the older bridge's data to classify the newer bridge's state. The result rests on a favourable source–target pair: two structures of the same design on the same site. It shows that convergent evidence is achievable, but a single case cannot support an estimate of field reliability, and it says little about transfer between structurally dissimilar bridges.

Sensitivity analysis. The headline estimates were recomputed under ten alternative assumptions (Table 3; details in Supplementary Table S4). In every scenario, D4, M3, and the R4+A4+G2/G3 combination remain at or below 0.8% of the corpus, G2/G3 remains below 4.5%, and at least 44% of studies demonstrate none of the trustworthiness attributes. Two scenarios move the estimates most. First-pass codes raise G2/G3 from 16 to 33, because the strict transfer rule was applied only in the final pass. Counting synthetic damage on real measurements (A3) as authentic raises R4+A4 from 6 to 56, which shows how much the conclusion depends on keeping A3 separate. Restricting the analysis to the original 727 studies changes the high-maturity counts by at most one study, except T2 (20 to 16), because several recall-audit additions analyse a year or more of monitoring data. Replacing

the codes of the 52 full-text-verified studies leaves D4, G2/G3, and M3 unchanged and raises R4+A4 from six to seven, through a study of an in-service bridge with documented cracking that had been coded as numerical (Table S6).

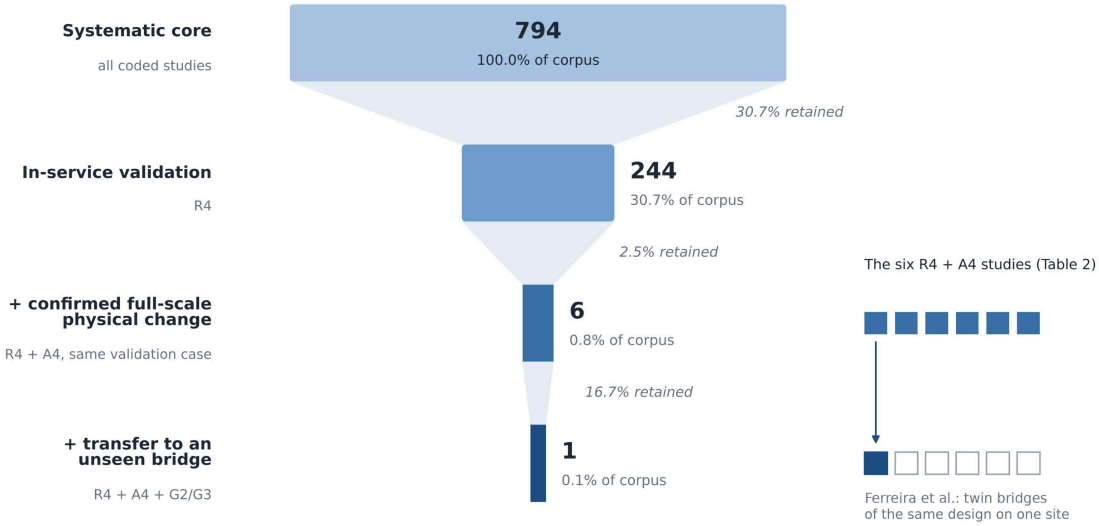


Figure 6. Evidence-convergence funnel (n = 794). Of 244 studies with in-service validation (30.7% of the corpus), six also demonstrate confirmed full-scale physical change in the same validation case (2.5% of the R4 tier; 0.8% of the corpus), and one of these, the twin-bridge study of Ferreira et al. [12], also demonstrates transfer to an unseen bridge (16.7% of the R4+A4 tier; 0.1% of the corpus). Tier widths are proportional to the number of studies, with a minimum width for the two smallest tiers; the squares on the right show the six R4+A4 studies (Table 2) and the one that also transfers. The final tier shows that convergence is feasible; it is not an estimate of field reliability. The six blue squares depict the R4+A4 studies listed in Table 2; the single filled dark-blue square identifies the twin-bridge study by Ferreira et al. [12], while outlined squares denote the other five studies.

Table 3. Sensitivity of the headline estimates to coding assumptions, corpus composition, and benchmark reuse.

Scenario	D4	R4+A4	G2/G3	T2	M3	No Attribute (%)	R4+A4+G2/G3
Submitted ledger (n = 727)	6	8	15	16	3	67.0	1
Revised baseline (n = 794)	5	6	16	20	1	65.2	1
Original 727 only, corrected	4	6	15	16	0	67.0	1
First-pass codes	4	3	33	21	1	44.3	0
Never-recoded studies only (n = 396)	1	0	5	7	1	63.9	0
Excluding uncertain additions	4	6	15	20	0	65.4	1
A3 counted as A4	5	56	16	20	1	65.2	1
T2 requires $\geq$ 2 years	5	6	16	10	1	65.2	1
T2 from $\geq$ 6 months	5	6	16	22	1	65.2	1
One publication per named bridge (n = 713)	5	6	12	18	1	65.1	1
Full-text-verified codes for 52 studies	5	7	16	21	1	64.7	1

Source: Supplementary Table S4. Revised baseline = values reported in Sections 4–8. D4 and M3 values reflect the full-text recoding of two studies [59,103] (Table S3a); the submitted-ledger, first-pass, and never-recoded rows are unaffected.

Benchmark reuse. Publication counts are not counts of independent validation events. In total, 102 studies (12.8%) use one of 21 named bridges or benchmark datasets, including 40 that use Z24, 11 Old ADA, and 11 KW51. The effect is largest for damage authenticity:

52 of the 61 A4 studies use a named benchmark, and counting each named bridge once reduces A4 from 61 to 13 publications. Coding consistency was also checked across studies of the same bridge. All Z24 and Old ADA studies carry the same R3/A4 codes; mixed codes for KW51, Yonghe, and I-40 reflect studies that use FE models of these bridges instead of their measured data. Two inconsistencies were found and corrected: a Powder Mill Bridge study was recoded from A4 to A3, because its damage was simulated in a calibrated FE model [108], and an S101 study was recoded from R4 to R3, because the S101 tests were a dedicated controlled campaign [64].

9.2. Trustworthiness Trade-Offs and Boundary Cases

The five attributes are complementary, not additive. Physics integration excludes implausible solutions but can transfer model-form error from uncertain material properties, boundary conditions, or reduced-order assumptions [2,24]. Calibrated uncertainty on one bridge does not establish transfer [95]. An attribution shows which measurements drive a decision but is not proof of a structural cause [53]. Fusion adds synchronisation, covariance, and missing-data dependencies [56,128], and environmental normalisation reduces false alarms but can suppress damage-sensitive information [8,9]. The rarity of the U+X pair (two studies) and of four-attribute studies (1) shows that these interactions have seldom been evaluated together.

The low first-pass stability of E, U, and F arose from recurring boundary cases. Gaussian noise added to simulated signals had been coded as E. Bayesian updating or probabilistic wording with point outputs had been coded as U, whereas only studies reporting posterior or predictive distributions for the inferred quantity, such as [95,117], qualify. Serial hybrids such as CNN–long short-term memory (LSTM) networks had been coded as F, whereas fusion requires separate information streams, as in multi-domain [13] or decision-level fusion [56]. Supplementary Table S3 lists each recoded study. Because these attributes are the most interpretation-sensitive, their prevalence values are conservative lower bounds.

9.3. Practical Implications for Bridge Owners

The framework can be used as an evidence-screening checklist, not as a certification score. Before adopting an AI-supported monitoring method, an owner can put seven questions to the evidence behind it:

- Was the method validated numerically, in a controlled full-scale campaign, or on an operating bridge? (R)
- Was real physical deterioration present, or was the damage simulated or injected? (A).
- Were temperature, traffic, wind, and other operating factors addressed? (E)
- Did the method work on a bridge other than the one it was developed on, and how much data from that bridge did it need? (G)
- Is the uncertainty of the output quantified, and can the basis of a warning be explained? (U, X)
- How long was the evaluation period? (T)
- Is the output linked to a defined inspection, intervention, or maintenance step? (M)

The corpus sets a baseline for these questions: fewer than 1 in 100 published studies validate a method on an operating bridge with confirmed damage, about 1 in 50 demonstrates transfer to another bridge, and about two in three report none of the trustworthiness attributes. In risk-based schemes such as the Italian guidelines for existing bridges, which assign attention classes and set inspection and monitoring requirements accordingly [129], an M1 output can inform the condition record, an M2 output can trigger a special inspection or a change of attention class, and only an M3 output with stated uncertainty can

inform intervention planning. Regulatory acceptance, procurement, liability, cybersecurity, and acceptable-risk thresholds lie outside the M scale. The evidence reviewed supports a human-in-the-loop workflow in which an AI alert is followed by engineering review, targeted inspection or conventional NDT, and a structural decision.

9.4. Limitations

First, the 794 records are publications, not independent validation events, and benchmark reuse inflates the apparent breadth of the evidence, especially for A4. Second, publication-level codes compress heterogeneous experiments, which is why combinations are counted only within a validation case. Third, the corpus is restricted to English-language journal articles, so null results, proprietary industrial systems, and unpublished operational experience are under-represented. Publication bias would tend to overstate capability and would therefore not weaken the finding that convergent evidence is scarce, but undocumented field deployments could raise the R4, T2, or M2–M3 counts. Fourth, attributes are coded from what is reported, so a missing code means that the evidence was not found. Fifth, agreement between the original title/abstract decisions and an independent human screener was moderate ($\kappa = 0.43$), and the 80 records that the human screener advanced but the original screening excluded were not assessed at full text, so eligible studies may be missing from the corpus. Sixth, the coding-stability and full-text verification statistics measure consistency and accuracy against the source documents, not inter-rater reliability between independent human coders; the verification indicates that M2 is over-coded and D3, T1, and T2 are under-coded, and the 67 recall-audit additions were coded from abstracts. Seventh, the bridge-primary boundary excludes generic-structure studies that use a bridge as one of several test cases. Finally, T2 measures the span of evaluated data, not long-term deployment, and the review covers only vibration and dynamic-response AI; GPR [5], computer vision, ultrasonic and impact-echo testing, thermography, and acoustic emission lie outside the quantitative denominator.

10. Conclusions

AI-based vibration assessment of bridges has demonstrated broad diagnostic capability, but the evidence needed for deployment has not kept pace. Of 794 coded studies, only six (0.8%) validate a method in service against confirmed physical change, and only one of these, for two nearly identical neighbouring bridges, also transfers to an unseen bridge. The specific findings are as follows:

(1) Diagnostic capability is concentrated in enabling inference and detection (D0 and D1: 67.8%); prognosis of structural condition (D4) is demonstrated in five studies (0.6%) and must be distinguished from response forecasting.
(2) Validation realism and damage authenticity are independent. Numerical validation is the largest category (45.5%), and in-service validation covers 30.7%, but confirmed full-scale change appears in 61 studies, 40 of them on Z24, and in 13 when each named bridge is counted once.
(3) Verified transfer to an unseen bridge or bridge population is demonstrated in 16 studies (2.0%), five of them in service.
(4) Environmental robustness is demonstrated in 21.8% of studies, and uncertainty quantification, explainability, physics integration, and fusion each in 6.0% or fewer; 65.2% of studies demonstrate none of the five attributes, and uncertainty and explainability occur together in two studies.
(5) Year-scale evaluation (T2, 20 studies), prognosis (D4, 5), and maintenance decision support (M3, 1) are rare, and none of the 13 classifiable T2 studies operated a fixed model prospectively.

These findings hold across ten sensitivity scenarios, including first-pass codes, the stable-coded subset, alternative authenticity and temporal thresholds, benchmark-collapsed counts, and full-text-verified codes.

Five research directions follow from these gaps; each tied to the dimensions it would raise:

(i) Prospective validation of locked models on in-service bridges with independently documented deterioration (R4, A4, T2);
(ii) Transfer tests on held-out, structurally dissimilar bridges, with stated target-data budgets and tests for negative transfer (G2/G3);
(iii) Calibrated uncertainty and domain-shift detection that allow a model to abstain outside its validated range, evaluated jointly with explainability and physics integration so that the trade-offs between trustworthiness attributes are measured (U, E, X, P);
(iv) Longitudinal prognosis verified against later inspection and linked to a defined maintenance decision (D4, M3);
(v) Shared multi-bridge datasets with documented damage provenance and time-separated blind evaluation, together with reporting standards that make the D/R/A/G/T/M evidence of each new study explicit.

The framework used here gives a reproducible way to tell methodological promise from engineering validation, and to plan studies around the evidence that deployment requires.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ndt4040033/s1.

Author Contributions: Conceptualization, M.Z.B., M.S. and M.L.P.; methodology, M.Z.B.; formal analysis, M.Z.B.; investigation, M.Z.B.; data curation, M.Z.B.; writing—original draft preparation, M.Z.B.; writing—review and editing, M.S. and M.L.P.; visualization, M.Z.B. supervision, M.S. and M.L.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study is available on request from the corresponding author.

Acknowledgments: ChatGPT 4.8 (OpenAI) and Claude Opus 5.5 (Anthropic) were used to assist eligibility assessment, and coding. The authors checked all outputs, revised and proofread the text, and take full responsibility for the content of the article.

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

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
CNN	Convolutional neural network
D0–D4	Diagnostic capability (enabling inference to prognosis)
R1–R4	Validation realism (numerical to in-service)
A0–A4	Damage authenticity (none to confirmed full-scale physical change)
G0–G3	Generalisation (same structure to bridge population)
T0–T2	Temporal maturity (evaluated data span)
M0–M3	Engineering-use maturity (algorithmic output to decision support)
E, U, X, P, F	Trustworthiness attributes: environmental and operational robustness, uncertainty quantification, explainability, physics integration, information fusion

ZS, UT, FS, ST	Target supervision: zero-shot, unsupervised target information, few-shot labelled target data, supervised retraining
EOV	Environmental and operational variability
FE	Finite element
GPR	Ground-penetrating radar
InSAR	Interferometric synthetic aperture radar
LSTM	Long short-term memory
ML	Machine learning
NDT	Nondestructive testing
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RMSE	Root-mean-square error
SHM	Structural health monitoring
UQ	Uncertainty quantification
VBI	Vehicle–bridge interaction
XAI	Explainable artificial intelligence

References

1. Di Mucci, V.M.; Cardellicchio, A.; Ruggieri, S.; Nettis, A.; Renò, V.; Uva, G. Artificial intelligence in structural health management of existing bridges. *Autom. Constr.* **2024**, *167*, 105719. [CrossRef]
2. Mammeri, S.; Barros, B.; Conde-Carnero, B.; Riveiro, B. From traditional damage detection methods to Physics-Informed Machine Learning in bridges: A review. *Eng. Struct.* **2025**, *330*, 119862. [CrossRef]
3. Niyirora, R.; Ji, W.; Masengesho, E.; Munyaneza, J.; Niyonyungu, F.; Nyirandayisabye, R. Intelligent damage diagnosis in bridges using vibration-based monitoring approaches and machine learning: A systematic review. *Results Eng.* **2022**, *16*, 100761. [CrossRef]
4. Bao, Y.; Sun, H.; Xu, Y.; Guan, X.; Pan, Q.; Liu, D. Recent advances in structural health diagnosis: A machine learning perspective. *Adv. Bridge Eng.* **2025**, *6*, 7. [CrossRef]
5. Yu, Y.; Rashidi, M.; Dorafshan, S.; Samali, B.; Noroozinejad Farsangi, E.; Yi, S.; Ding, Z. Ground penetrating radar-based automated defect identification of bridge decks: A hybrid approach. *J. Civ. Struct. Health Monit.* **2025**, *15*, 521–543. [CrossRef]
6. Yu, Y.; Li, J.; Li, J.; Xia, Y.; Ding, Z.; Samali, B. Automated damage diagnosis of concrete jack arch beam using optimized deep stacked autoencoders and multi-sensor fusion. *Dev. Built Environ.* **2023**, *14*, 100128. [CrossRef]
7. Corbally, R.; Malekjafarian, A. A data-driven approach for drive-by damage detection in bridges considering the influence of temperature change. *Eng. Struct.* **2022**, *253*, 113783. [CrossRef]
8. Pei, X.-Y.; Zhang, H.-T.; Huang, H.-B.; Liang, D. Probabilistic Machine Learning-Based Frequency Normalization Method for Bridge Damage Detection Considering Environmental Variations. *Int. J. Struct. Stab. Dyn.* **2026**, *26*, 26501099. [CrossRef]
9. Ahmad, H.; Matsumoto, Y. Damage detection in a single-span prestressed concrete girder bridge under environmental variations using Gaussian process regression validated by physics-guided surrogate modelling. *J. Civ. Struct. Health Monit.* **2026**, *16*, 50. [CrossRef]
10. Bacha, M.Z.; Puppio, M.L.; Zucca, M.; Sassu, M. Dynamic Response-Based Bridge Monitoring and Structural Assessment: A Structured Scoping Review and Evidence Inventory. *Infrastructures* **2026**, *11*, 134. [CrossRef]
11. Duran, B.; Azam, Y.E.; Linzell, D.G. Bridge health monitoring via drive-by sensing: A novel CNN-based framework with full-scale field validation. *Mech. Syst. Signal Process.* **2026**, *252*, 114255. [CrossRef]
12. Ferreira, L.; Yano, M.O.; Souza, L.; Moldovan, I.; da Silva, S.; Lopes, R.; Cimini, C.A.; Costa, J.C.W.A.; Figueiredo, E. Transfer learning and Bayesian calibration addressing data scarcity and uncertainty for structural health monitoring of twin concrete bridges. *Mech. Syst. Signal Process.* **2025**, *235*, 112845. [CrossRef]
13. Mousavi, V.; Rashidi, M.; Ghazimoghadam, S.; Samali, B. A data-driven digital twin model for bridge health monitoring using feature fusion and unsupervised deep learning. *Inf. Fusion* **2026**, *126*, 103534. [CrossRef]
14. Jasiński, M.; Fawad, M.; Sabzi Khoshraftar, A.; Abbozzo, A.; Tao, Y.; Salamak, M.; Kopeć, B.; Chen, Q. Model-based predictive digital twin for bridge structural health monitoring: The integration of building information modelling, finite element analysis, and machine learning in a Netherlands case study. *Eng. Appl. Artif. Intell.* **2026**, *181*, 115458. [CrossRef]
15. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]

16. Malekjafarian, A.; Golpayegani, F.; Moloney, C.; Clarke, S. A machine learning approach to bridge-damage detection using responses measured on a passing vehicle. *Sensors* **2019**, *19*, 4035. [CrossRef] [PubMed]
17. Bragança, C.S.C.D.; Souza, E.; Ames, I.; Ribeiro, D.; Bittencourt, T.N. AI-based drive-by methodology for early damage detection in a steel truss railway bridge under multiple environmental and operational factors. *Struct. Infrastruct. Eng.* **2026**, *online first*. [CrossRef]
18. Khan, A.Q.; Raza, A.; Pimanmas, A. Physics-informed neural networks for early-warning fatigue damage detection and localisation in a steel railway bridge using sparse ambient monitoring data. *Intell. Transp. Infrastruct.* **2026**, *5*, liag010. [CrossRef]
19. Chen, C.; Huang, X.; Liu, Z.; Yang, B.; Zhou, L.; Jiang, Z.; Liu, Y.; Tang, L. Long-term structural health monitoring strain modelling for bridges via data correlation and deep learning. *Eng. Res. Express* **2025**, *7*, 025125. [CrossRef]
20. Deng, P.-H.; Cui, C.; Cheng, Z.-Y.; Zhang, Q.-H.; Bu, Y.-Z. Fatigue damage prognosis of orthotropic steel deck based on data-driven LSTM. *J. Constr. Steel Res.* **2023**, *202*, 107777. [CrossRef]
21. Wang, Z.; Zhan, J.; Ni, Z.; Xu, X.; Wu, L.; Wang, C. A physics-enhanced deep learning framework for baseline-free damage localisation in railway continuous beam bridge using displacement influence line. *Adv. Eng. Inform.* **2026**, *76*, 105039. [CrossRef]
22. Wan, H.-P.; Fang, T.-L.; Zhu, Y.-K.; Wang, C.; Wang, N.-B. An improved SIFT-based method for non-contact bridge displacement measurement. *Adv. Eng. Inform.* **2026**, *71*, 104145. [CrossRef]
23. Ye, X.; Chen, X.; Lei, Y.; Fan, J.; Mei, L. An integrated machine learning algorithm for separating the long-term deflection data of prestressed concrete bridges. *Sensors* **2018**, *18*, 4070. [CrossRef] [PubMed]
24. Cianci, E.; Civera, M.; De Biagi, V.; Chiaia, B. Physics-informed machine learning for the structural health monitoring and early warning of a long highway viaduct with displacement transducers. *Mech. Syst. Signal Process.* **2026**, *242*, 113659. [CrossRef]
25. Yang, X.-M.; Zhang, Z.-K.; Chen, T.; Wang, Z.; Yao, X.-J. A transformer-based framework for automating frequency domain decomposition in bridge modal analysis. *Structures* **2026**, *88*, 111912. [CrossRef]
26. Mu, H.-Q.; Zheng, Z.-J.; Wu, X.-H.; Su, C. Bayesian network-based modal frequency-multiple environmental factors pattern recognition for the Xinguang Bridge using long-term monitoring data. *J. Low Freq. Noise Vib. Act. Control* **2020**, *39*, 545–559. [CrossRef]
27. Mao, J.-X.; Wang, H.; Spencer, B.F. Gaussian mixture model for automated tracking of modal parameters of long-span bridge. *Smart Struct. Syst.* **2019**, *24*, 243–256. [CrossRef]
28. Calderon Hurtado, A.; Xu, J.; Salleh, R.; Dias-da-Costa, D.; Alamdari, M.M. Development and field validation of a fully customised vehicle scanning system on two full-scale bridges. *Struct. Health Monit.* **2026**, *online first*. [CrossRef]
29. Fernandes, T.M.; Minski, L.; De Souza, P.V.G.; Ribeiro, D.R.F.; Miguel, L.F.F.; Lopez, R.H. Early Scour Damage Detection Using Drive-By Monitoring Data through Supervised Learning. *J. Struct. Des. Constr. Pract.* **2026**, *31*, 04026031. [CrossRef]
30. Nayek, R.; Narasimhan, S. Extraction of contact-point response in indirect bridge health monitoring using an input estimation approach. *J. Civ. Struct. Health Monit.* **2020**, *10*, 815–831. [CrossRef]
31. Tyler, G.; Hurtado, A.C.; Hamedani, S.J.; Salleh, R.; Alamdari, M.M. Isolation distributional kernels for indirect bridge health monitoring: A full-scale investigation of anomaly detection. *Comput.-Aided Civ. Infrastruct. Eng.* **2026**, *50*, 100099. [CrossRef]
32. Nguyen, D.H.; Nguyen, Q.B.; Bui-Tien, T.; De Roeck, G.; Abdel Wahab, M. Damage detection in girder bridges using modal curvatures gapped smoothing method and Convolutional Neural Network: Application to Bo Nghi bridge. *Theor. Appl. Fract. Mech.* **2020**, *109*, 102728. [CrossRef]
33. Dackermann, U.; Smith, W.A.; Alamdari, M.M.; Li, J.; Randall, R.B. Cepstrum-based damage identification in structures with progressive damage. *Struct. Health Monit.* **2019**, *18*, 87–102. [CrossRef]
34. Huang, S.; Zeng, W.; Deng, X.; Yang, Y.B. Bridge Damage Detection in Vehicle-Bridge Interaction System Under Road Surface Roughness: Convolutional Neural Networks with Baseline-Referenced Residual Transformation. *Int. J. Struct. Stab. Dyn.* **2026**, *26*, 27503329. [CrossRef]
35. Guo, J.; Zhang, G. Deep learning based identification of vortex-induced vibration in stay cables using multidimensional feature fusion and BiLSTM-MHA. *J. Wind Eng. Ind. Aerodyn.* **2026**, *273*, 106424. [CrossRef]
36. Mehrjoo, A.; Hom, K.L.; Beigi, H.; Betti, R. Zero-Shot Bridge Health Monitoring Using Cepstral Features and Streaming LSTM Networks. *Infrastructures* **2025**, *10*, 292. [CrossRef]
37. Nnamani, N.E.; Matos, J.C.; Komarizadehasl, S.; Nguyen, N.T.T.; Dang, S.N. A Hybrid CNN-LSTM Framework for Vibration-Based Multi-Damage Assessment in Reinforced Concrete Bridges. *Appl. Sci.* **2026**, *16*, 6659. [CrossRef]
38. Khodabandehlou, H.; Pekcan, G.; Fadali, M.S. Vibration-based structural condition assessment using convolution neural networks. *Struct. Control Health Monit.* **2019**, *26*, e2308. [CrossRef]
39. Avcı, M.S.; Ercan, E.; Hızal, Ç.; Malekjafarian, A.; Ozer, E. Cross-structure domain translation with structural state awareness: AT-StarGAN-GP for multi-level synthetic signal generation in structural health monitoring. *Struct. Health Monit.* **2026**, *online first*. [CrossRef]
40. Lu, N.; Xiao, X.; Cui, J.; Liu, Y.; Huang, K.; Yuen, K.-V. An unsupervised cross-domain method for bridge damage detection based on multichannel symmetric dot pattern feature alignment. *Comput.-Aided Civ. Infrastruct. Eng.* **2025**, *40*, 5698–5718. [CrossRef]

41. Yano, M.O.; Figueiredo, E.; da Silva, S.; Cury, A. Foundations and applicability of transfer learning for structural health monitoring of bridges. *Mech. Syst. Signal Process.* **2023**, *204*, 110766. [CrossRef]
42. Al-Adly, A.I.F.; Kripakaran, P. Developing physics-informed neural networks for virtual sensing in beams with moving loads. *Eng. Struct.* **2025**, *338*, 120535. [CrossRef]
43. Abu Zouriq, M.F.; Linzell, D.G.; Azam, Y.E. Integrating Bi-Bi-LSTM-Based Virtual Sensing and VAE for Sensor-Efficient Unsupervised Structural Damage Detection. *J. Eng. Mech.* **2026**, *152*, 04026037. [CrossRef]
44. Yang, Y.; Feng, D.; Xu, Y. A TCN-PatchTST framework for efficient reconstruction of missing acceleration data in bridge structural health monitoring. *Eng. Struct.* **2026**, *366*, 123548. [CrossRef]
45. Tian, H.; Zhu, Q.; Zhang, Q.; Wang, X.; Du, Y. A time-frequency heterogeneous dual-domain joint sampling-reconstruction and uncertainty quantification network for structural monitoring data reconstruction. *Struct. Health Monit.* **2026**, *online first*. [CrossRef]
46. Xiong, J.; Hu, L.; Meng, X.; An, X.; Xie, Y. Cross-Modal Graph Attention for Bridge SHM Data Imputation. *Sensors* **2026**, *26*, 3339. [CrossRef]
47. Zhang, M.; Guo, T.; Zhang, G.; Liu, Z.; Liu, Y. Missing monitoring data reconstruction for cable-stayed bridge using knowledge transfer-based generative pre-trained model. *Struct. Health Monit.* **2025**, *online first*. [CrossRef]
48. Longji, Z.; Zhi, Y.; Jiaqing, L.; Wenhua, L.; Jingchun, M. Spatiotemporal dependency data imputation for long-term health monitoring of concrete arch bridges. *Sci. Rep.* **2025**, *15*, 36218. [CrossRef] [PubMed]
49. Wu, D.; Li, X.; Xu, N.; Li, Y.; Jin, Y.; He, S.; Xu, W.; Li, N.; Chen, W.; Laima, S. Cross-attention based frequency-domain prediction method for buffeting response of long-span bridges. *Adv. Bridge Eng.* **2026**, *7*, 36. [CrossRef]
50. Chen, R.; Kim, C.-W.; Wang, J. Physics-informed neural operator for forecasting vehicle-induced bridge vibration trajectories. *Eng. Appl. Artif. Intell.* **2026**, *176*, 114708. [CrossRef]
51. Li, Z.; Yan, B.; Meng, Q.; Xu, C.; Zhang, F.; Wang, Y.; Domingo, M. Strain prediction in a large-span arch bridge using the TimeXer model considering temperature and traffic loads. *Front. Built Environ.* **2026**, *12*, 1749222. [CrossRef]
52. Yang, Y.-L.; Zhu, Y.-C.; Cai, C.S.; Wu, S.-H. Fusion of Bayesian Time Domain and Gaussian Process Models for Modal Identification Under Environmental Variations. *Struct. Control Health Monit.* **2026**, *2026*, 3260954. [CrossRef]
53. Haghbin, M.; Tomassini, E.; Ubertini, F.; García-Macías, E.; Chiachío-Ruano, J. Explainable AI for Operational Modal Analysis: Field deployment on densely instrumented structures. *Eng. Struct.* **2026**, *358*, 122631. [CrossRef]
54. Ahmad, H.; Matsumoto, Y. Environmental effect compensation and anomaly detection in an ageing prestressed concrete girder bridge using Gaussian process regression. *Struct. Infrastruct. Eng.* **2026**, *online first*. [CrossRef]
55. Colacillo, P.; Civera, M.; Surace, C.; Tronci, E.M. Integrating climate projections with Gaussian processes for structural anomaly detection via cointegration. *J. Low Freq. Noise Vib. Act. Control* **2026**, *online first*. [CrossRef]
56. Teng, S.; Chen, G.; Liu, Z.; Cheng, L.; Sun, X. Multi-sensor and decision-level fusion-based structural damage detection using a one-dimensional convolutional neural network. *Sensors* **2021**, *21*, 3950. [CrossRef] [PubMed]
57. Qu, G.; Song, M.; Xia, Y.; Sun, L. Bridge Girder-End Displacement Reconstruction Using a Novel Hybrid Attention Mechanism Leveraging Multisource Information. *Struct. Control Health Monit.* **2025**, *2025*, 8249455. [CrossRef]
58. Zhou, B.; Zhu, J.; Jiao, X.; Yessoufou, F. Damage localisation based on fusion of multi-source vehicle-induced response of bridges. *Struct. Infrastruct. Eng.* **2026**, *online first*. [CrossRef]
59. Hakimi, O.; Liu, H.; Abudayyeh, O. Deep learning-driven multi-level data fusion framework for predictive maintenance and structural health monitoring of concrete bridge decks. *Autom. Constr.* **2025**, *175*, 106180. [CrossRef]
60. Chun, P.-J.; Yamashita, H.; Furukawa, S. Bridge Damage Severity Quantification Using Multipoint Acceleration Measurement and Artificial Neural Networks. *Shock Vib.* **2015**, *2015*, 789384. [CrossRef]
61. Tan, Z.X.; Thambiratnam, D.P.; Chan, T.H.T.; Gordan, M.; Abdul Razak, H. Damage detection in steel-concrete composite bridge using vibration characteristics and artificial neural network. *Struct. Infrastruct. Eng.* **2020**, *16*, 1247–1261. [CrossRef]
62. Fallahian, M.; Khoshnoudian, F.; Meruane, V. Ensemble classification method for structural damage assessment under varying temperature. *Struct. Health Monit.* **2018**, *17*, 747–762. [CrossRef]
63. Xiang, C.; Zhao, H.; Wu, G.; Chen, L.; Yang, Z.; Patelli, E. Damage diagnosis of bridge structures using deep learning strategies: A hybrid neural networks practical tool. *Struct. Health Monit.* **2026**, *online first*. [CrossRef]
64. Das, T.; Guchhait, S. Sequential optimisation architecture for bridge damage localisation and severity estimation under dynamic moving loads using an automated machine learning-driven hybrid deep learning model. *Eng. Appl. Artif. Intell.* **2026**, *177*, 114968. [CrossRef]
65. Impraimakis, M.; Palkanoglou, E.N. A generative adversarial network optimisation method for damage detection and digital twinning by deep AI fault learning: Z24 Bridge structural health monitoring benchmark validation. *Struct. Multidiscip. Optim.* **2025**, *68*, 221. [CrossRef]
66. Qiu, Y.; Ahmed, B.; Abueidda, D.W.; El-Sekelly, W.; de Soto, B.G.; Abdoun, T.; Ji, H.; Qiu, J.; Mobasher, M.E. Damage identification for bridges using machine learning: Development and application to KW51 bridge. *Digit. Eng.* **2026**, *10*, 100103. [CrossRef]

67. Riyahi, A.; Mestari, M.; Bouihi, B. Data-Driven and Physics-Informed Neural Networks for Structural Health Monitoring of the Z24 Bridge. *J. Civ. Eng. Forum* **2026**, *16*, 24173. [CrossRef]
68. Hadizadeh, A.; Tarighat, A.; Malian, A. A data-driven framework for structural health monitoring using reinforcement learning and deep autoencoders. *Sci. Rep.* **2026**, *16*, 23157. [CrossRef] [PubMed]
69. Lee, K.; Hwang, J.; Shin, D.H.; Lee, J.-H. A time-interval-based incremental learning paradigm for progressive bridge damage detection using convolutional autoencoders. *Struct. Infrastruct. Eng.* **2026**, *online first*. [CrossRef]
70. Giorgi, V.; Tordela, C.; Bernardini, L.; Ramírez Balbiano, P.A.; Somaschini, C.; Strano, S.; Terzo, M. An LSTM Autoencoder-Based Approach for Monitoring Railway Bridges. *Appl. Sci.* **2026**, *16*, 4272. [CrossRef]
71. Giglioni, V.; Venanzi, I.; Poggioni, V.; Milani, A.; Ubertini, F. Autoencoders for unsupervised real-time bridge health assessment. *Comput.-Aided Civ. Infrastruct. Eng.* **2023**, *38*, 959–974. [CrossRef]
72. Nesackon Abraham, J.; Tran, M.Q.; Jayaraj, J.S.; Matos, J.C.; Valluzzi, M.R.; Dang, S.N. Unsupervised Learning-Based Anomaly Detection for Bridge Structural Health Monitoring: Identifying Deviations from Normal Structural Behaviour. *Sensors* **2026**, *26*, 561. [CrossRef] [PubMed]
73. Gonzalez, I.; Karoumi, R. BWIM aided damage detection in bridges using machine learning. *J. Civ. Struct. Health Monit.* **2015**, *5*, 715–725. [CrossRef]
74. Ni, Y.-Q.; Wang, J.; Chan, T.H.T. Structural damage alarming and localisation of cable-supported bridges using multi-novelty indices: A feasibility study. *Struct. Eng. Mech.* **2015**, *54*, 337–362. [CrossRef]
75. Arangio, S.; Bontempi, F. Structural health monitoring of a cable-stayed bridge with Bayesian neural networks. *Struct. Infrastruct. Eng.* **2015**, *11*, 575–587. [CrossRef]
76. Park, S.; Kim, S. Hybrid CAE-DSVDD for unsupervised vibration-based damage detection in in situ steel truss bridge. *Struct. Health Monit.* **2025**, *online first*. [CrossRef]
77. Bayane, I.; Leander, J.; Karoumi, R. An unsupervised machine learning approach for real-time damage detection in bridges. *Eng. Struct.* **2024**, *308*, 117971. [CrossRef]
78. Sarwar, M.Z.; Cantero, D. Probabilistic autoencoder-based bridge damage assessment using train-induced responses. *Mech. Syst. Signal Process.* **2024**, *208*, 111046. [CrossRef]
79. Yang, J.; Liu, D.; Zhao, L.; Yang, X.; Li, R.; Jiang, S.; Li, J. Improved stochastic configuration network for bridge damage and anomaly detection using long-term monitoring data. *Inf. Sci.* **2025**, *700*, 121831. [CrossRef]
80. Shi, S.; Du, D.; Mercan, O.; Kalkan, E.; Parol, J. A novel decentralized damage detection method for self-powered wireless sensing in structural health monitoring using self-supervised learning. *Eng. Struct.* **2025**, *344*, 121399. [CrossRef]
81. Shi, S.; Du, D.; Mercan, O.; Kalkan, E.; Parol, J. Contrastive and self-supervised learning for open-set damage classification in structural health monitoring with incomplete and imbalanced vibration data. *Expert Syst. Appl.* **2025**, *293*, 128731. [CrossRef]
82. Liu, J.; Zhang, W.; Sun, L.; Li, Y. Physics-encoded unsupervised deep learning for lightweight structural health monitoring of short and medium-span bridge networks. *Eng. Appl. Artif. Intell.* **2026**, *181*, 115515. [CrossRef]
83. Heravi, M.A.; Soleimani-Babakamali, M.H.; Naderpour, H.; Sadhu, A. Zero-shot transfer learning for structural damage detection using target-to-source structure domain data mapping. *Mech. Syst. Signal Process.* **2026**, *242*, 113673. [CrossRef]
84. Wan, C.; Hou, J.; Zhang, G.; Gao, S.; Ding, Y.; Cao, S.; Hu, H.; Xue, S. Domain adaptation based automatic identification method of vortex induced vibration of long-span bridges without prior information. *Eng. Appl. Artif. Intell.* **2025**, *139*, 109677. [CrossRef]
85. Liu, J.; Xu, S.; Bergés, M.; Noh, H.Y. HierMUD: Hierarchical multi-task unsupervised domain adaptation between bridges for drive-by damage diagnosis. *Struct. Health Monit.* **2023**, *22*, 1941–1968. [CrossRef]
86. Hou, J.; Cao, S.; Hu, H.; Zhou, Z.; Wan, C.; Noori, M.; Li, P.; Luo, Y. Vortex-Induced Vibration Recognition for Long-Span Bridges Based on Transfer Component Analysis. *Buildings* **2023**, *13*, 2012. [CrossRef]
87. Hu, Z.; He, W.; Li, H.; Wu, Y. Few-Shot Cross-Bridge Damage Diagnosis from Vibration Sensor Signals via Siamese Contrastive Pretraining with Self-Calibrated Convolution. *Sensors* **2026**, *26*, 4153. [CrossRef] [PubMed]
88. Lu, N.; Cui, J.; Zeng, W.; Xiao, X.; Luo, Y. Damage identification of a reduced-scale cable-stayed bridge based on domain adaptation transfer learning. *Meas. J. Int. Meas. Confed.* **2026**, *257*, 118674. [CrossRef]
89. Giglioni, V.; Poole, J.; Mills, R.; Venanzi, I.; Ubertini, F.; Worden, K. Transfer learning in bridge monitoring: Laboratory study on domain adaptation for population-based SHM of multispan continuous girder bridges. *Mech. Syst. Signal Process.* **2025**, *224*, 112151. [CrossRef]
90. Giglioni, V.; Eva, A.E.; Worden, K.; Ubertini, F.; Venanzi, I. Assessing similarity requirements for effective transfer learning across a network of rigid frame bridges. *Reliab. Eng. Syst. Saf.* **2026**, *268*, 111994. [CrossRef]
91. Ma, S.; Flanigan, K.A.; Bergés, M. Bridging the Reality Gap in Digital Twins with Context-Aware: Physics-Guided Deep Learning. *J. Comput. Civ. Eng.* **2026**, *40*, 04026009. [CrossRef]
92. Wu, S.-H.; Zhu, Y.-C.; Pan, Z.-H.; Di, H. An efficient Expectation-Maximization algorithm for Bayesian operational modal analysis with physics-data fusion model. *Mech. Syst. Signal Process.* **2025**, *237*, 113144. [CrossRef]

93. Zeng, J.; Cao, Z.-J.; Wan, Q.; Chen, R.; Chen, H. Generative diffusion-aided probabilistic spatiotemporal response reconstruction for structural dynamic system under sparse noisy measurements. *Comput.-Aided Civ. Infrastruct. Eng.* **2026**, *50*, 100108. [CrossRef]
94. Yu, H.; Shu, J.; Dong, T.; Yang, H.; Chen, Q.; Wang, J.; Guo, S. Missing structural response recovery under different damage scenarios via a multi-modal diffusion model fusing inspection text and monitoring data. *Adv. Eng. Inform.* **2026**, *76*, 105113. [CrossRef]
95. Jiang, W.-J.; Kim, C.-W. Ambient vibration-based cable tension monitoring and uncertainty analysis in cable-stayed bridge from a fully probabilistic perspective. *Eng. Struct.* **2025**, *328*, 119755. [CrossRef]
96. Pan, C.; Gong, F.; Liu, X.; Yan, B.; Xia, Y. Axle-information-free identification of vehicle moving forces from bridge displacement using U-Net. *Adv. Struct. Eng.* **2026**, *online first*. [CrossRef]
97. Wu, Y.; Deng, L.; He, W. Bwimnet: A novel method for identifying moving vehicles utilizing a modified encoder-decoder architecture. *Sensors* **2020**, *20*, 7170. [CrossRef] [PubMed]
98. Ying, L.Q.; Ying, G.G.; Hu, J.L.; Zhang, W.D. Deep Learning Method for Vehicle Load Identification Using Digital Twin and Transfer Learning Strategies. *J. Bridge Eng.* **2026**, *31*. [CrossRef]
99. Yang, H.; Yan, W.; He, H. Parameters identification of moving load using ANN and dynamic strain. *Shock Vib.* **2016**, *2016*, 8249851. [CrossRef]
100. Fan, L.; Guo, Z.; Shi, H. A Bayesian-Boosting Method for Probability Distribution Estimation of Structural Damping. *Struct. Control Health Monit.* **2026**, *2026*, 2872248. [CrossRef]
101. Nguyen, T.Q.; Nguyen, D.P.; Nguyen, P.T.; Nguyen, T.T. Bayesian Deep Learning for Bridge Structural Damage Assessment Based on Dissipation Coefficient Evaluation. *Int. J. Steel Struct.* **2026**, *online first*. [CrossRef]
102. Xie, C.; Bai, Y. A physics-informed neural network for predicting structural fatigue damage of orthotropic bridge deck through updating model uncertainties. *Int. J. Fatigue* **2026**, *202*, 109253. [CrossRef]
103. Cao, K.; Ding, Y.; Li, H.; Yang, T.-L.; Liu, G.-G.; Chu, T.-Y. Bridge cables vibration frequency identification and fatigue life prediction based on SHM data with machine learning. *Steel Compos. Struct.* **2025**, *55*, 451–468. [CrossRef]
104. Lu, Q.; Zhu, J.; Zhang, W. Quantification of Fatigue Damage for Structural Details in Slender Coastal Bridges Using Machine Learning-Based Methods. *J. Bridge Eng.* **2020**, *25*, 04020033. [CrossRef]
105. Wang, T.; Li, P.; Zou, S.; Zhao, D.; Wang, X.; Guo, K. Corrosion fatigue analysis on in-service steel bridges subject to random traffic loads via long short-term memory-based surrogate model. *Structures* **2025**, *76*, 108890. [CrossRef]
106. Hom, K.L.; Beigi, H.; Betti, R. From laboratory experiments to in-service tests: Learning embeddings for damage identification in structural health monitoring. *Mech. Syst. Signal Process.* **2026**, *242*, 113622. [CrossRef]
107. Weinstein, J.C.; Sanayei, M.; Brenner, B.R. Bridge Damage Identification Using Artificial Neural Networks. *J. Bridge Eng.* **2018**, *23*, 0001302. [CrossRef]
108. Kaspar, K.; Santini-Bell, E.; Petrik, M.; Sanayei, M. Comparison between a Linear Regression and an Artificial Neural Network Model to Detect and Localize Damage in the Powder Mill Bridge. *Transp. Res. Rec.* **2020**, *2674*, 394–404. [CrossRef]
109. Li, S.; Wang, W.; Lu, B.; Du, X.; Dong, M.; Zhang, T.; Bai, Z. Long-term structural health monitoring for bridge based on back propagation neural network and long and short-term memory. *Struct. Health Monit.* **2023**, *22*, 2325–2345. [CrossRef]
110. Deng, P.; Yang, J.J.; Yee, T.; Oguzmert, M. Regime-Aware Bridge Scour Depth Estimation Using Mixture of Experts and Accelerometer Data. *Transp. Res. Rec.* **2026**, *online first*. [CrossRef]
111. Jin, C.; Jang, S.; Sun, X.; Li, J.; Christenson, R. Damage detection of a highway bridge under severe temperature changes using extended Kalman filter trained neural network. *J. Civ. Struct. Health Monit.* **2016**, *6*, 545–560. [CrossRef]
112. Park, H.S.; Hong, T.; Lee, D.-E.; Oh, B.K.; Glisic, B. Long-term structural response prediction models for concrete structures using weather data, fiber-optic sensing, and convolutional neural network. *Expert Syst. Appl.* **2022**, *201*, 117152. [CrossRef]
113. Giglioni, V.; Poole, J.; Venanzi, I.; Ubertini, F.; Worden, K. A domain adaptation approach to damage classification with an application to bridge monitoring. *Mech. Syst. Signal Process.* **2024**, *209*, 111135. [CrossRef]
114. Duran, B.; Eftekhar Azam, S.; Sanayei, M. Leveraging Deep Learning for Robust Structural Damage Detection and Classification: A Transfer Learning Approach via CNN. *Infrastructures* **2024**, *9*, 229. [CrossRef]
115. Sadeghi Eshkevari, S.; Marasco, G.; Sen, D.; Dabbaghchian, I.; Pakzad, S.N. Assessing uncertainty in CNN-attention-based indirect bridge strain estimation from field acceleration data. *Struct. Infrastruct. Eng.* **2026**, *22*, 812–826. [CrossRef]
116. Asadollahi, P.; Huang, Y.; Li, J. Bayesian finite element model updating and assessment of cable-stayed bridges using wireless sensor data. *Sensors* **2018**, *18*, 3057. [CrossRef] [PubMed]
117. Gong, F.; Xia, Y.; Ling, Z.; Lozano, F.; He, T. Bayesian deep learning based bridge condition assessment considering uncertainty quantification of missing data. *Eng. Struct.* **2026**, *348*, 121753. [CrossRef]
118. Deng, P.; Yang, J.J.; Yee, T.; Oguzmert, M. A Physics-Guided Feature Fusion Network for Vibration-Based Bridge Scour Monitoring. *Transp. Res. Rec.* **2026**, *online first*. [CrossRef]
119. Kaewnuratchadasorn, C.; Wang, J.; Kim, C.-W.; Yang, Y. Two-Dimensional Vehicle-Bridge Interaction Neural Operator for Digital Twin of Bridge Structures. *Struct. Control Health Monit.* **2025**, *2025*, 1031214. [CrossRef]

120. Armijo, A.; Zamora-Sánchez, D. Integration of Railway Bridge Structural Health Monitoring into the Internet of Things with a Digital Twin: A Case Study. *Sensors* **2024**, *24*, 2115. [CrossRef] [PubMed]
121. Park, S.; Chang, M.; Yazdanpanah, O. A condition monitoring of steel box girder bridges based on consecutive system identification and model updating using a long-term monitoring database. *Structures* **2025**, *80*, 109779. [CrossRef]
122. Al-Hijazeen, A.; Koris, K. Feedforward Neural Network-Based Digital Twin for SHM of Bridges. *Archit. Civ. Eng. Environ.* **2025**, *18*, 157–169. [CrossRef]
123. Lei, Z.; Zhu, L.; Fang, Y.; Niu, C.; Zhao, Y. Fiber Bragg Grating Smart Material and Structural Health Monitoring System Based on Digital Twin Drive. *J. Nanomater.* **2022**, *2022*, 4356974. [CrossRef]
124. Yan, F.; Geng, Y.; Liu, Y.; Chen, N.; Li, X.; Wang, Y.; Aloisio, A. A vortex-induced vibration warning method based on ensemble-learning-embedded neural network. *Comput.-Aided Civ. Infrastruct. Eng.* **2025**, *40*, 6419–6436. [CrossRef]
125. Wu, W.-H.; Chen, C.-C.; Chen, Z.-T.; Lai, G. Automated real-time cable tension warning framework using long-term monitoring data and a convolutional neural network. *Struct. Health Monit.* **2026**, *online first*. [CrossRef]
126. Ge, L.; Wang, C.; Liu, M.; Dan, D.; Jian, F. VIV-SDE-Net: A physics-informed neural framework for long-horizon prediction of bridge vortex-induced vibrations. *Adv. Struct. Eng.* **2026**, *online first*. [CrossRef]
127. Sánchez-Haro, J.; García, M.; Capellán, G.; da Costa, A.; Perez, P.; Añó, J. Digital twin for predictive maintenance on the Espartxo Bridge: Application to early detection of under-foundation scour. *Structures* **2025**, *71*, 107916. [CrossRef]
128. Hassani, S.; Dackermann, U.; Mousavi, M.; Mustapha, S.; Li, J. AAE-CycleWGAN fusion framework for generating fused strain data from sparse to dense domains in bridge monitoring systems. *Inf. Fusion* **2026**, *127*, 103736. [CrossRef]
129. Ministero delle Infrastrutture e della Mobilità Sostenibili. *Linee Guida per la Classificazione e Gestione del Rischio, la Valutazione della Sicurezza ed il Monitoraggio dei Ponti Esistenti [Guidelines for the Classification and Management of Risk, Safety Assessment and Monitoring of Existing Bridges]*; Decreto Ministeriale 1 Luglio 2022, n. 204; Gazzetta Ufficiale della Repubblica Italiana, Serie Generale n. 196, 23 August 2022. First Adopted by Decreto Ministeriale 17 Dicembre 2020, n. 578; Ministero delle Infrastrutture e della Mobilità Sostenibili: Roma, Italy, 2020.

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.