infrastructures

MDPI

Article

Synthetic Data Generation for the Prototyping of Bridge Damage Detection Algorithms

Matvei Sinden and Alejandro Jiménez Rios *

Department of Architecture & Civil Engineering, University of Bath, Bath BA2 7AY, UK
* Correspondence: ajr225@bath.ac.uk

Highlights

- Calibrated S101 FE model used to generate 7300 synthetic modal frequency iterations.
- Thermal frequency shifts up to 1.33% masked minor structural damage signals.
- Supervised neural networks outperformed k-means clustering on thermally variant data.
- Linear FE modelling failed to replicate experimental pier settlement response.
- FAIR-compliant synthetic dataset provided for benchmarking SHM algorithms.

Abstract

The application of Machine Learning (ML) to bridge Structural Health Monitoring (SHM) is constrained by a lack of diverse and labelled datasets. Obtaining high-quality training data from operational infrastructure is inherently difficult because critical assets are typically repaired immediately upon the detection of defects, preventing the collection of data describing diverse failure modes. To address this scarcity and enable the prototyping of robust algorithms, this study presents a framework for generating synthetic modal frequencies using a calibrated Finite Element (FE) model of the S101 bridge. Aleatory uncertainties and environmental variability are incorporated through the stochastic variation of material properties and thermal loads derived from a 20-year climate record. Analysis of the generated dataset revealed that simulated thermal loads induced frequency shifts that often exceeded those caused by minor structural damage, confirming the necessity of training on environmentally representative data. The primary contribution of this work is an open-access, FAIR-compliant (Findable, Accessible, Interoperable, Reusable) synthetic dataset, intended to serve as a standardised benchmark for the SHM research community under conditions of combined structural and environmental uncertainty. To demonstrate the utility of the generated data, the performance of a supervised multi-layer perceptron and an unsupervised k-means clustering algorithm are evaluated, with the supervised approach achieving a maximum classification accuracy of 1.00. However, the framework also reveals a fundamental modelling limitation: the linear FE approach failed to replicate the physical response under pier settlement, producing frequency shifts an order of magnitude below those observed experimentally.

Keywords: machine learning; S101 bridge; thermal masking; digital twin; finite element modelling; structural health monitoring; anomaly identification

Check for updates

Academic Editors: Mi G. Chorzepa and Armando La Scala

Received: 9 June 2026
Revised: 1 August 2026
Accepted: 7 August 2026
Published: 21 August 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

The long-term structural reliability of transport infrastructure is increasingly challenged by the degradation of ageing assets, environmental exposure and rising operational

demands [1,2]. Traditional maintenance strategies have predominantly relied on visual inspections, yet these methods are labour-intensive, subjective and often limited to detecting surface-level defects in accessible areas [3]. To overcome these limitations and ensure safety, Structural Health Monitoring (SHM) has transitioned towards continuous, global monitoring systems. Vibration-based methods, which depend on the principle that changes in physical properties (specifically stiffness and mass) manifest as detectable shifts in modal parameters (natural frequencies, mode shapes and damping ratios), have become a cornerstone of this approach [4–6].

Recent advancements in Machine Learning (ML) have offered new opportunities to automate the interpretation of these complex vibration datasets, moving the field from manual data analysis to automated decision-making [7,8]. The current state of the art can be broadly categorised into supervised and unsupervised learning approaches. Studies have demonstrated that supervised learning algorithms, such as Neural Networks (NNs), are capable of identifying non-linear patterns indicative of damage with high accuracy, provided they are trained on comprehensive datasets [9–11]. Unsupervised methods, including clustering algorithms and outlier analysis, have also shown promise in identifying structural deviations (novelty detection) without requiring exhaustive prior knowledge of damage states. The interpretation of these deviations, however, often requires specialised expert judgement to disentangle genuine damage features from environmental noise, a process that necessitates careful attention to minimise false positives under operational variance [12,13]. More recent hybrid architectures, combining convolutional layers with bidirectional gated recurrent units and attention mechanisms, suggest that robustness to environmental temperature variation can instead be built directly into the network itself rather than treated purely as a downstream classification problem [14].

The widespread adoption of these data-driven techniques in civil engineering remains constrained by a critical data scarcity gap [15]. Unlike in other domains such as computer vision, obtaining large, labelled datasets from real-world bridges in damaged states is difficult in practice. Critical infrastructure is typically repaired immediately upon the detection of significant defects to ensure public safety, preventing the collection of data describing diverse failure modes [16,17]. Therefore, most available real-world datasets only represent the "healthy" or "reference" state of the structure, rendering high-performance ML classifiers ineffective for practical deployment.

Compounding the issue of data scarcity is the influence of environmental and operational variability. It is well-established in the literature that fluctuations in ambient temperature can induce shifts in natural frequencies that may mask or even exceed the shifts caused by structural damage [18–21]. Temperature typically causes a more pronounced effect than that of wind and precipitation [22,23]. As a result, algorithms trained on limited data often struggle to distinguish between benign environmental changes and genuine structural deterioration, particularly as there is no agreed-upon relationship between temperature and modal frequency response [24,25]. Training on temperature-variant datasets improves the distinction between damage and environmental effects [26], with methods such as Principal Component Analysis (PCA) having also demonstrated success in separating thermal variation from genuine structural change [27]. Comparable hybrid frameworks, coupling support vector machines with metaheuristic optimisation, have been used to quantify ambient temperature directly from natural frequencies before locating and sizing damage, with good accuracy reported even under simulated measurement noise [28]. Generating synthetic data with applied thermal loads is therefore vital for robust algorithm training.

Physics-based frameworks utilising modal analysis provide a robust means of generating diverse damage scenarios grounded in structural mechanics [29], with representative

numerical models identified as a necessity to overcome the practical impossibility of emulating damage on operational assets to obtain diverse training data [9]. By calibrating numerical models against empirical benchmarks, such as the destructive testing campaign of the S101 post-tensioned concrete bridge by Siringoringo et al. [30], it is possible to simulate a vast array of damage scenarios that are unfeasible to capture in the field. However, it is noted that physics-based models often rely on simplified assumptions, such as linear elastic behaviour of materials, which lack the variability inherent in real-world structures or fail to adequately integrate long-term climatic patterns [31]. Furthermore, despite the existence of feature extraction benchmarks [32], there is a distinct absence of open-access, FAIR-compliant (Findable, Accessible, Interoperable, Reusable) datasets that allow for the rigorous benchmarking of ML algorithms under conditions of combined damage and environmental uncertainty [16,33]. Whilst recent standardisation efforts, such as BS EN 18162:2026 [34], have begun to formalise the concept of digital twins in the built environment, the practical implementation of such frameworks for SHM remains constrained by the absence of representative datasets that capture realistic damage and environmental conditions.

In summary, bridge infrastructure faces mounting safety challenges, as evidenced both by the empirical record of structural failures [35] and by the escalating threats of climate change, whose short-term hazards such as flooding and scour interact with long-term stressors such as corrosion and thermal degradation to compromise bridge safety at scale [2]. This is particularly relevant for historic bridges that are currently subjected to comparatively higher loads than those they were originally designed and built to support [36]. These pressures have intensified the demand for robust SHM and damage detection frameworks capable of supporting proactive asset management. Physics-enhanced machine learning and digital twinning have emerged as particularly promising paradigms in this regard [37]. However, a critical bottleneck remains: real-world monitoring data is expensive to acquire, limited in volume, and rarely captures the full spectrum of damage states relevant for algorithm development.

The present paper addresses this gap, proposing a FAIR-compliant synthetic data generation methodology specifically tailored to the prototyping of bridge damage detection algorithms, thereby advancing the field towards more generalisable and practically deployable SHM solutions. This is achieved through the simultaneous integration of three sources of variability absent from prior work on this structure: stochastic aleatory uncertainty in material properties and boundary conditions, long-term thermal loading derived from a 20-year historical climate record [38] and a range of representative structural damage scenarios. Furthermore, a quantified demonstration of thermal masking effects across multiple damage typologies and severities is conducted, providing direct evidence to inform the design of environmentally compensated ML classifiers. To demonstrate the utility of the generated data, the performances of a supervised NN (multi-layer perceptron) [9] and an unsupervised k-means clustering algorithm [12] are evaluated to assess the suitability of the synthetic database for classifying damage states amidst simulated environmental noise.

2. Materials and Methods

This study employs a physics-based framework to generate a comprehensive synthetic dataset suitable for training ML algorithms. The methodology leverages a calibrated Finite Element (FE) model of the S101 bridge to simulate the structure's dynamic response under diverse conditions, comprising stochastic aleatory uncertainties, realistic thermal loads and distinct damage scenarios. The resulting dataset was subsequently used to train and evaluate the performance of both supervised and unsupervised ML classifiers.

The S101 bridge was a post-tensioned prestressed concrete flyover located on the A1 Westautobahn in Austria and constructed during the early 1960s. The bridge was originally designed as a three-span structure comprising a 32 m main span and two 12 m side spans, resulting in an overall length of 58 m. The deck is a continuous 7.2 m wide element and consists of a double-webbed T-beam cross-section, with webs 0.6 m thick. The beam depth varies from 0.9 m at mid-span to 1.7 m above the piers, reflecting the variable-depth design commonly adopted for prestressed concrete bridges of that period (see Figure 1). The bridge was ultimately scheduled for demolition because its load-carrying capacity and maintenance condition no longer satisfied contemporary requirements, while its geometry was also incompatible with the evolving transport infrastructure. Prior to demolition, the bridge became the subject of an extensive experimental campaign, making it one of the best-documented full-scale benchmark structures for structural health monitoring research. Photographs of the bridge can be consulted elsewhere [30,39].

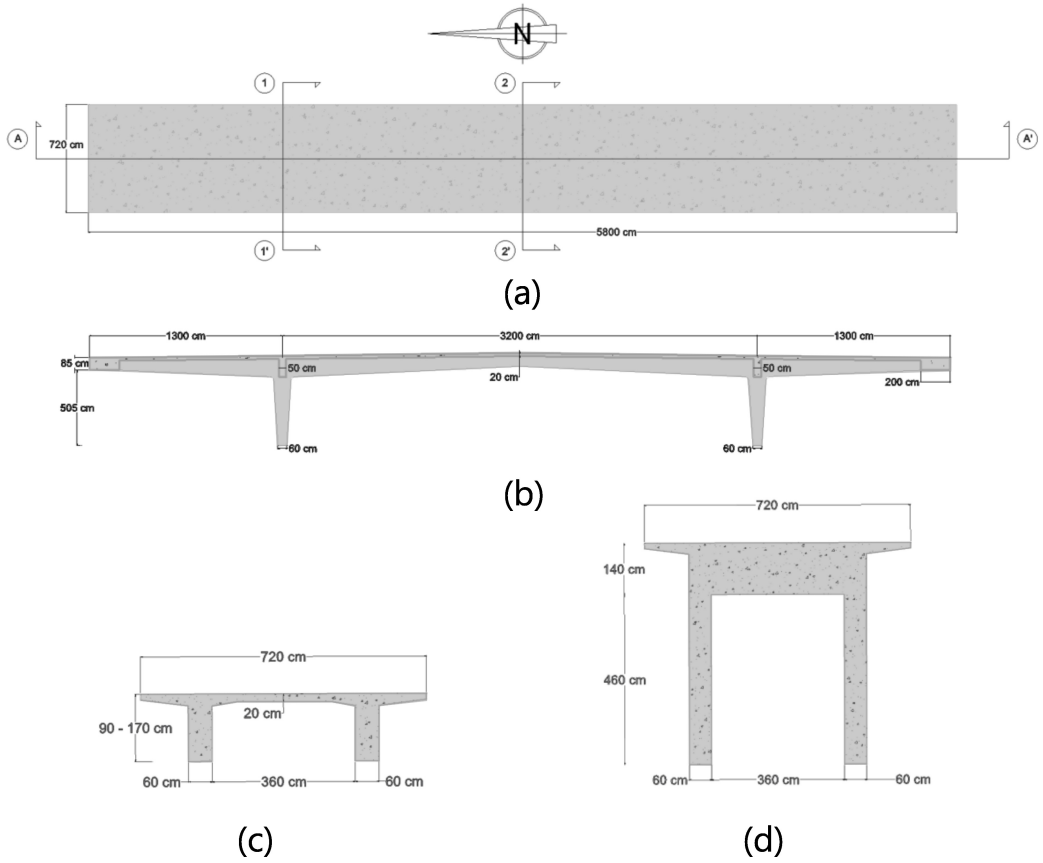

Figure 1. Geometry of the S101 bridge, a three-span continuous rigid-frame bridge, showing its (**a**) plan view, and three sections: (**b**) section A-A', (**c**) section 2-2', and (**d**) section 1-1', and . Not to scale.

2.1. Finite Element Model Setup and Uncertainty Simulation

The synthetic data generation utilised a calibrated FE model of the S101 bridge developed using Ansys, as detailed by Jiménez Rios et al. [40]. This model serves as the baseline for the undamaged condition, having been calibrated against empirical data from progressive damage testing [30] to ensure physical fidelity.

The FE model was developed in Ansys Workbench (version 2024 R2) by first importing the bridge geometry from an IGES file into the Geometry module. A Static Structural analysis system was subsequently configured to define the material properties, mesh, boundary conditions, applied loads and solver settings. A Modal analysis system was then incorporated into the workflow, inheriting the solution from the static analysis to determine

the natural frequencies of the structure under self-weight. The bridge was subjected to its self-weight, assuming a concrete density of 2400 kg/m³ and a gravitational acceleration of 9.81 m/s², together with an additional uniformly distributed mass of 583 kg/m² applied along the side surfaces of the deck to represent the weight of the sidewalks and metallic safety rails. Contact interfaces between structural components were automatically detected and modelled as bonded contacts. The numerical mesh was generated using a patch-conforming algorithm with higher-order ten-node tetrahedral SOLID187 elements, each possessing three translational degrees of freedom. A mesh sensitivity analysis was carried out using element sizes ranging from 1.0 m to 0.1 m, from which an element size of 0.5 m was selected as providing the most appropriate balance between computational efficiency and numerical accuracy.

To produce a representative dataset that accounts for real-world variability, aleatory uncertainties were introduced into the model. The random perturbations introduced in the finite element model are intended to emulate the small variability commonly observed in experimentally identified modal properties due to sensor precision, data acquisition systems and signal processing, rather than to reproduce the intrinsic variability of the structural material properties themselves [41,42]. In practice, repeated measurements of the same bridge under nominally identical structural, environmental and loading conditions seldom produce identical modal parameters because of unavoidable measurement uncertainty. Within the proposed framework, this effect is approximated by applying small random perturbations (2% coefficient of variation) to selected model parameters, thereby generating realistic scatter in the simulated modal frequencies without altering the underlying structural state. This approach provides a practical representation of measurement-induced variability commonly encountered in vibration-based structural health monitoring while maintaining a computationally efficient synthetic data generation process.

A total of 365 independent realisations of the undamaged bridge were generated under both thermally loaded and non-thermal conditions, yielding two sets of 365 simulations for the baseline state. This number was chosen to correspond directly to the days of a calendar year, enabling a one-to-one mapping of historical daily mean temperature records to individual model iterations and ensuring a balanced sample size across simulation sets. Each realisation underwent a modal analysis to extract the first six natural frequencies, exceeding the four modes identified during the original S101 experimental campaign by Siringoringo et al. [30] and the five modes considered in the associated conference paper by Jiménez Rios et al. [40]. This encompasses the full set of experimentally characterised modes whilst providing additional higher-mode information to enrich the dataset.

2.2. Simulation of Damage Scenarios

The same simulation procedure was then applied to three distinct damage cases, each representing a realistic deterioration mechanism observed in ageing infrastructure. For each severity within all three damage cases, two equivalent sets of 365 thermally loaded and non-thermal realisations were generated. The damage scenarios were selected to broadly reflect the controlled pre-demolition testing campaign conducted on the S101 bridge [30], particularly pier weakening, and are summarised in Table 1.

The selected damage severities (10%, 20% and 40% stiffness reduction) were adopted to represent three progressively increasing levels of structural deterioration commonly considered in vibration-based SHM benchmark studies. Rather than corresponding to specific deterioration mechanisms or predefined structural limit states, these values were selected using engineering judgment to provide representative damage scenarios of increasing detectability. The 10% reduction represents an incipient damage condition whose influence on the modal properties may be partially obscured by the inherent variability associated

with measurement uncertainty and environmental effects. The 20% reduction represents a moderate damage state that robust damage detection algorithms should reliably identify, allowing asset managers to intervene before further deterioration compromises structural performance. Finally, the 40% reduction represents a severe damage scenario intended to provide a clear benchmark against which the performance of different machine learning algorithms can be evaluated. These representative severity levels therefore enable a systematic assessment of algorithm performance across a broad spectrum of realistic damage detectability conditions while maintaining a controlled and reproducible synthetic dataset.

Table 1. Defined damage cases and severities applied to the FE model (components depicted in Figure 2).

Damage Case	Mechanism	Model Component	Severities
Case 1	Pier Stiffness Reduction	Concrete Columns	10%, 20%, 40%
Case 2	Mid-Deck Stiffness Reduction	Concrete CD	10%, 20%, 40%
Case 3	Pier Settlement	Concrete Columns	3 cm, 5 cm, 10 cm

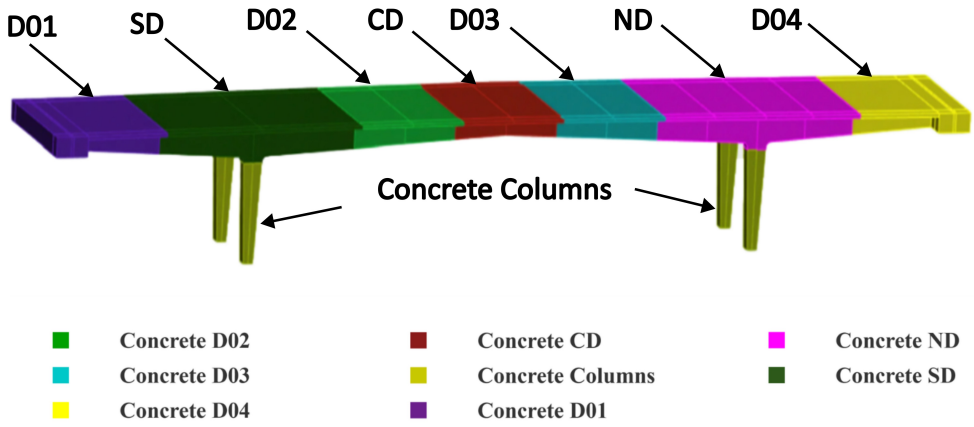

Figure 2. Calibrated FE model of the S101 bridge discretised by material component (the bridge deck spans 58 m in total length, with a main span of 32 m, adjacent spans of 12 m between centres and a width of 7.2 m [40]).

2.2.1. Stiffness Reduction (Cases 1 & 2)

Damage Cases 1 and 2 simulated material degradation, such as cracking or corrosion-induced internal voiding, through localised stiffness reductions [9]. These were implemented by reducing the Young's modulus of the affected component by 10%, 20% and 40%: a single *Concrete Columns* component (one pier) for Case 1 and a single *Concrete CD* component (one mid-span deck section) for Case 2 (Figure 2). To maintain consistency with the uncertainty modelling, the reduced stiffness values served as the new mean for the normal distribution, ensuring that stochastic variability remained present even within the damaged regions.

2.2.2. Pier Settlement (Case 3)

Case 3 simulated geometric deformation due to differential foundation settlement. Consistent with the experimental campaign on the S101 bridge [30], this was modelled by applying controlled vertical displacements of 3 cm, 5 cm and 10 cm to the base of a single *Concrete Columns* component (Figure 2). Boundary conditions were modified to permit vertical displacement while maintaining horizontal restraint.

2.3. Incorporation of Thermal Effects

To capture the influence of temperature fluctuations on modal frequencies, a steady-state thermal analysis was integrated into the simulation workflow. A dataset of 365 daily mean temperatures was compiled from a 20-year historical climate record for a nearby Austrian weather station [38], representing a full annual cycle, and applied uniformly to the structural model prior to each modal analysis. All undamaged and damaged simulations were run under these thermal loading conditions. A corresponding set of non-thermal simulations was also produced for each case, enabling direct assessment of the extent to which temperature fluctuations mask damage-induced frequency shifts.

It is acknowledged that this uniform steady-state approach represents a simplification of the complex spatial and temporal thermal gradients present in real structures; the implications of this assumption are discussed further in Section 3.4.

2.4. Automation and Data Handling

The generation of 7300 total model iterations (comprising 365 simulations across 20 distinct structural states) necessitated a fully automated workflow. Python (version 3.12) scripts were deployed within Ansys Workbench to iteratively assign stochastic material properties and boundary conditions based on a defined random seed, ensuring reproducibility. A secondary script within Ansys Mechanical managed the sequential execution of the steady-state thermal, static structural and modal analyses. The first six natural frequencies were extracted for each iteration and exported to CSV format.

Post-processing involved restructuring the simulation outputs into labelled datasets, with 0 and 1 denoting the undamaged and damaged states respectively, suitable for use in supervised learning. The datasets were compiled in accordance with FAIR principles [33] and are openly accessible via an online repository [43] to support future benchmarking and algorithm development.

2.5. Machine Learning Algorithm Development

Two distinct ML algorithms were developed to assess the suitability of the synthetically generated data for classifying the bridge's condition from modal frequency outputs: a supervised multi-layer perceptron NN and an unsupervised k-means clustering algorithm. For both algorithms, the dataset was partitioned into 70% training and 30% testing sets. A fixed random state was also applied to both algorithms to ensure the reproducibility of results.

2.5.1. Supervised Neural Network

A multi-layer perceptron classifier was implemented using the scikit-learn library. The architecture comprised an input layer of six neurons (corresponding to the first six natural frequencies), followed by four dense hidden layers containing 100, 50, 30 and 15 neurons, respectively. The ReLU activation function was applied to introduce non-linearity.

The model was optimised using the L-BFGS solver, with an alpha regularisation parameter of 1×10^{-3} to mitigate overfitting, and trained for up to 1000 iterations. Performance was evaluated using accuracy, the confusion matrix and the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC), alongside precision, recall, F1-score and balanced accuracy. To quantify statistical uncertainty and check for overfitting, each scenario was trained and evaluated over ten repeated runs with independently resampled train/test partitions, with the resulting mean, standard deviation, convergence behaviour and train–test performance gap reported in Appendix A.

The choice of hyperparameters, comprising the hidden-layer configuration and the alpha regularisation strength, was assessed through a five-fold stratified cross-validated

grid search applied independently to each of the eighteen damage scenarios. Candidate architectures ranged from a single 50-neuron layer up to the adopted four-layer network, and candidate regularisation strengths spanned 1×10^{-4} to 1×10^{-2}. The adopted regularisation strength of 1×10^{-3} was the most frequently selected value across the eighteen scenarios, supporting its use. For the architecture, simpler configurations achieved comparable, and in some scenarios marginally higher, cross-validated balanced accuracy than the adopted four-layer network. A single architecture was retained across all scenarios, both because it remained highly competitive in every case and to preserve a consistent model configuration for benchmarking purposes, rather than selecting a different architecture for each damage case. The full results are reported in Appendix A.

2.5.2. Unsupervised K-Means Clustering

To assess damage detection capabilities without labelled training data, a k-means clustering algorithm was employed. The algorithm partitioned the modal frequency data into two clusters (representing damaged and undamaged states) by minimising the distance to cluster centroids. The model was executed with 500 random initialisations and up to 1000 iterations to ensure stability. Performance was assessed using the Rand Index (RI), and PCA was utilised to reduce the dimensionality of data in order to visualise the separation of clusters in two-dimensional space.

2.5.3. Gaussian Mixture Model

As a further unsupervised benchmark, a Gaussian Mixture Model (GMM) with full covariance was implemented using the scikit-learn library. Two components were specified, matching the damaged and undamaged states, and the algorithm was initialised ten times to reduce sensitivity to local optima, with a maximum of 1000 expectation-maximisation iterations. Unlike k-means, which assumes spherical, equally-sized clusters, GMM models each component with its own covariance structure, allowing for more flexible cluster shapes. Performance was assessed using the same metrics and repeated-run procedure as the k-means algorithm.

3. Results and Discussion

This section presents the statistical distributions of the synthetically generated modal frequency data, evaluates the classification performance of the two ML algorithms and discusses the physical interpretation of the findings. The complete dataset [43] comprises 7300 model iterations, spanning 20 distinct structural states under varying thermal conditions, yielding 43,800 individual frequency datapoints in total.

Each set of model iterations comprises the first six natural frequencies of the bridge, corresponding to the mode shapes depicted in Figure 3. These encompass a mix of bending, torsional and lateral modes.

3.1. Influence of Environmental Variability

The application of thermal loads, derived from the 20-year historical climate record [38], produced a consistent reduction in mean natural frequencies across all modes relative to the non-thermal baseline. As illustrated in Figure 4, mean frequency shifts ranged from -0.13% to -1.33% across the six modes, with the largest shift observed in Mode 1 and the smallest in Mode 3. The interquartile range of the frequency distributions also increased notably under thermal loading, reflecting the additional variability introduced by seasonal temperature fluctuations. Scatter plots, illustrated in Figure 5, confirmed a negative translation of frequencies, particularly during simulated summer months, demonstrating the pronounced sensitivity of the bridge's dynamic response to ambient environmental conditions.

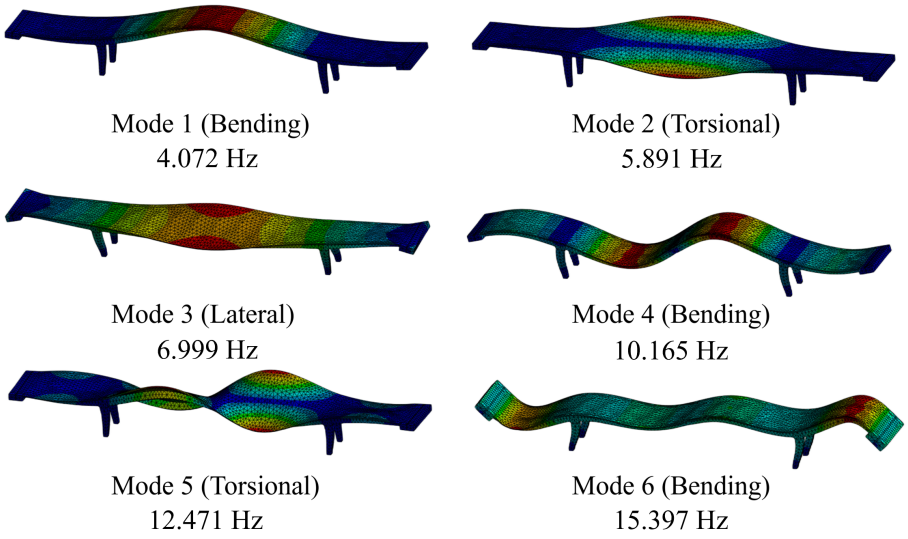

Figure 3. First six mode shapes (with respective frequencies) of the S101 bridge calibrated numerical model (not to scale).

These findings suggest that environmental variability poses a significant challenge to vibration-based SHM. Temperature-induced frequency shifts of up to −1.33% often exceeded those caused by minor structural damage, as demonstrated in the damage scenario analysis that follows. This appears consistent with observations by Figueiredo et al. [20] and Wang et al. [18], who noted that thermal effects can mask damage signals in operational structures. Consequently, the necessity of training algorithms on datasets that explicitly incorporate environmental variability is reinforced, as models trained solely on non-thermal data risk generating false positives under operational conditions [19].

3.2. Analysis of Damage Scenarios

3.2.1. Pier Stiffness Reduction (Case 1)

Simulated pier stiffness reductions produced detectable downward shifts in natural frequencies (Figure 6), the magnitude of which correlated consistently with damage severity. A 40% reduction in pier stiffness under thermal loads resulted in mean frequency shifts ranging from −0.51% to −1.27% across the six modes, with the largest shift observed in Mode 1 and the smallest in Mode 2 (Figure 7). Modes 1, 4 and 6 exhibited the greatest sensitivity. Mode 2 showed comparatively lower sensitivity at −0.51%, indicating limited modal participation from the affected pier. At lower severities, a 10% stiffness reduction produced shifts of approximately −0.36% for Mode 1, which resulted in substantial overlap between the damaged and undamaged frequency distributions once thermal variability was superimposed.

This overlap at low damage severities is particularly significant. The thermally induced mean frequency shift of −1.33% in Mode 1 exceeds the damage-induced shift of −0.36% at 10% pier stiffness reduction by a ratio of approximately 3.7, rendering the two states statistically indistinguishable without prior knowledge of the thermal contribution. Comparing the thermal and non-thermal simulations directly for this scenario, the addition of thermal loads changed the damage-induced frequency shifts by −0.02% (Mode 3) to −0.11% (Mode 1) across the six modes. In Mode 1, this supplementary thermal contribution is again nearly half as large as the −0.25% shift produced by the damage itself, demonstrating that thermal loading does not merely add background noise at this severity but introduces a shift of comparable magnitude to the damage signal itself. This behaviour illustrates precisely the challenge that motivated the inclusion of environmental variability

in the synthetic framework and shows why algorithms trained on non-thermal data are likely to underperform in operation. ML models trained on non-thermal data achieved perfect accuracy (1.00) for severe damage but saw performance drop markedly when thermal loads were introduced, as discussed further in Section 3.3.

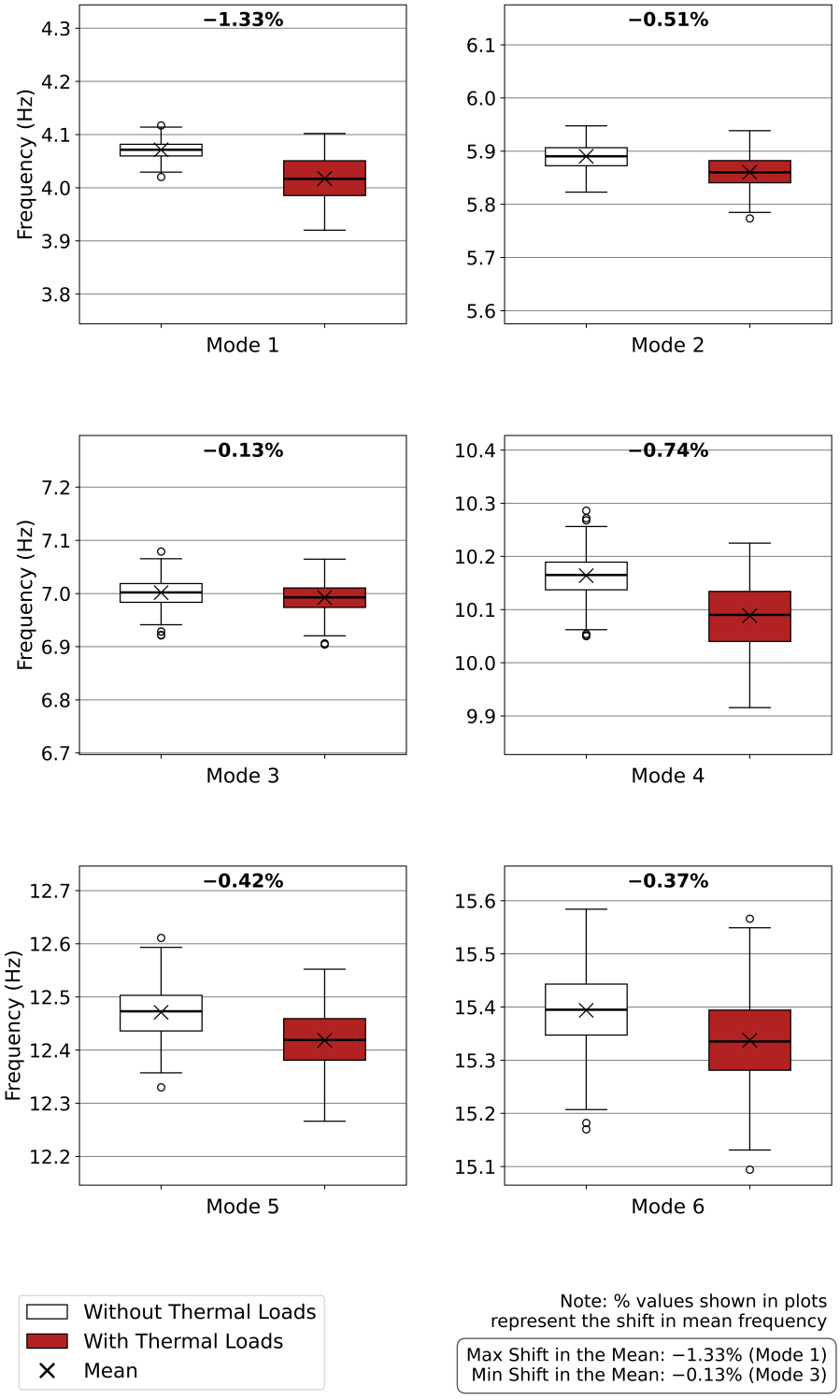

Figure 4. Box plots of natural frequency distributions for the undamaged case with and without thermal loads.

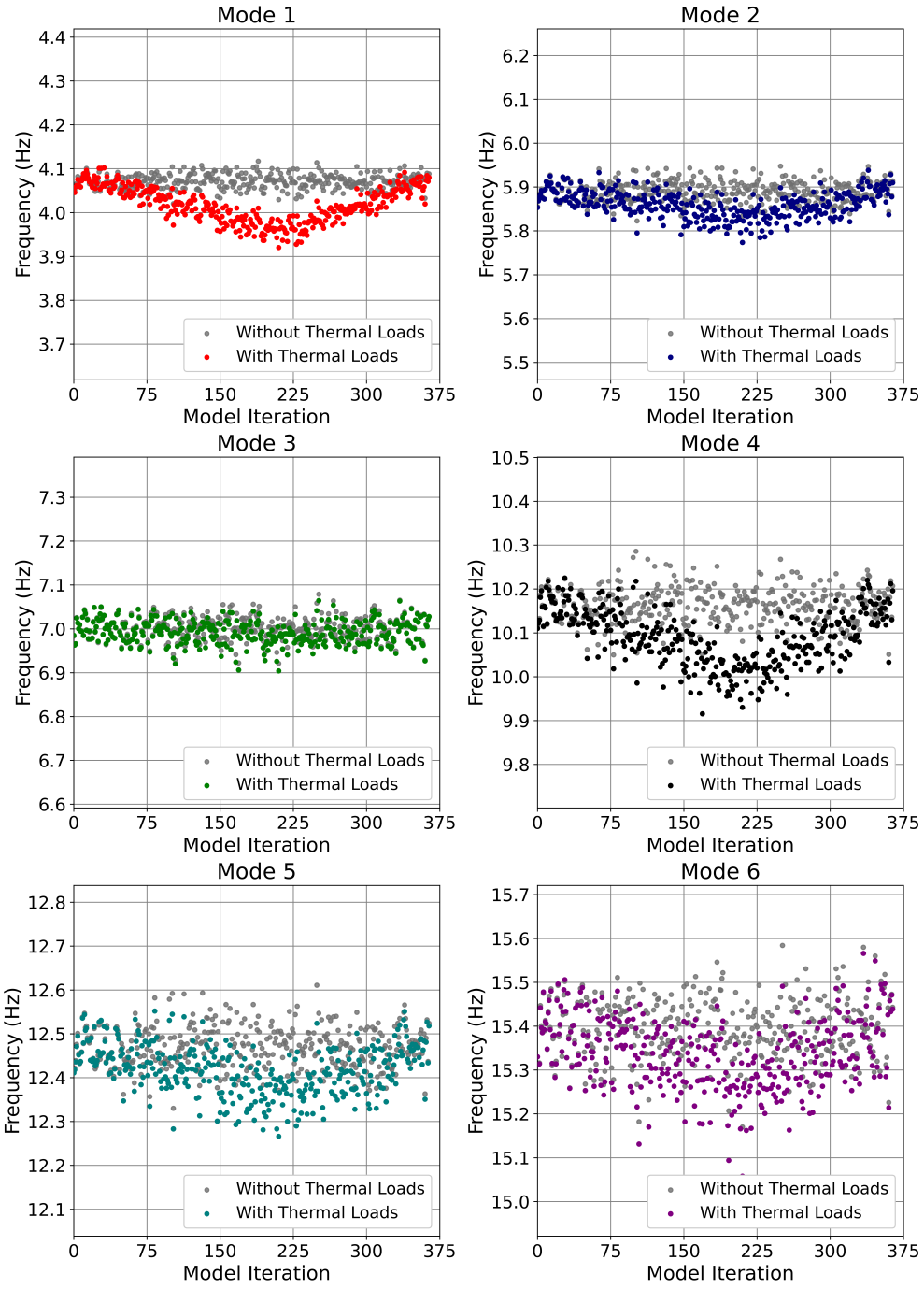

Figure 5. Scatter diagrams of natural frequency distributions for the undamaged case with and without thermal loads.

3.2.2. Mid-Deck Stiffness Reduction (Case 2)

Mid-deck stiffness reduction produced the most pronounced frequency deviations of all simulated damage cases. Under thermal loads, a 40% reduction in deck stiffness caused mean frequency shifts ranging from −0.24% to −3.35% across the six modes, with the largest shift observed in Mode 1 and the smallest in Mode 6 (Figure 8 and 9). Also notable was a −2.17% shift in Mode 5; these shifts were of sufficient magnitude to remain distinguishable from thermal noise across all tested severities. Mode 6, by contrast, exhibited negligible sensitivity at −0.24%, reflecting minimal participation of the mid-span section in that mode shape. Despite the imposed environmental variability, the separation between the damaged and undamaged distributions remained sufficient for reliable classification at 40% severity,

as reflected in the NN's mean accuracy of 1.00 under thermal loading, discussed further in Section 3.3.

The strong detectability of mid-deck damage across all severities, even under thermal loading, is attributable to the location and extent of the stiffness reduction. The mid-span region contributes substantially to the global bending and torsional modes of the deck, such that even a 10% reduction produces frequency shifts (−0.71% in Mode 1) that, whilst partially masked by temperature, retain sufficient separation from the undamaged distribution for the NN to classify reliably (0.90 accuracy). This contrasts distinctly with the pier stiffness case, where the localised nature of the damage limits its influence on the global modal response.

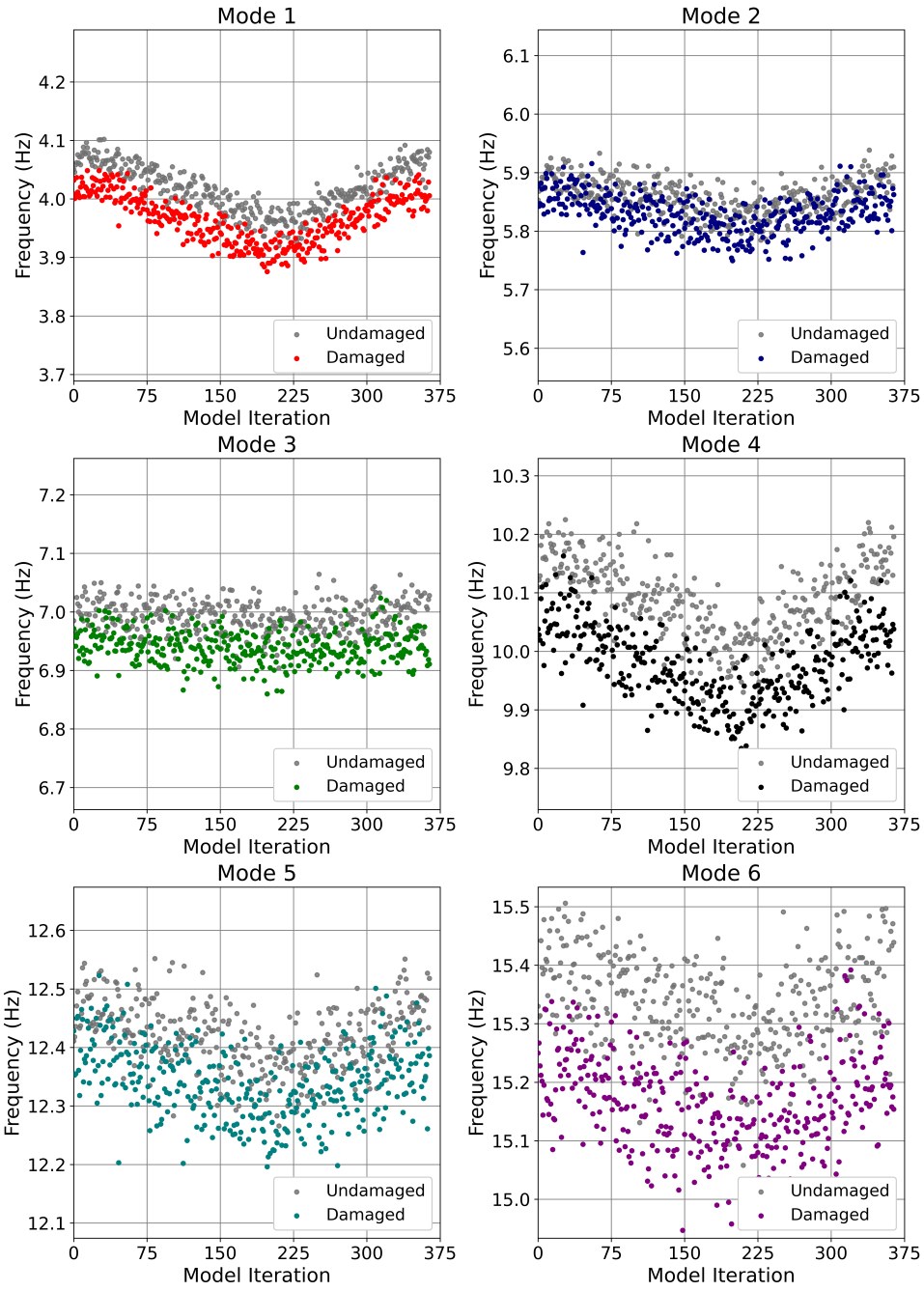

Figure 6. Scatter diagrams of natural frequency distributions for the undamaged and 40% pier stiffness reduction cases, both under thermal loads.

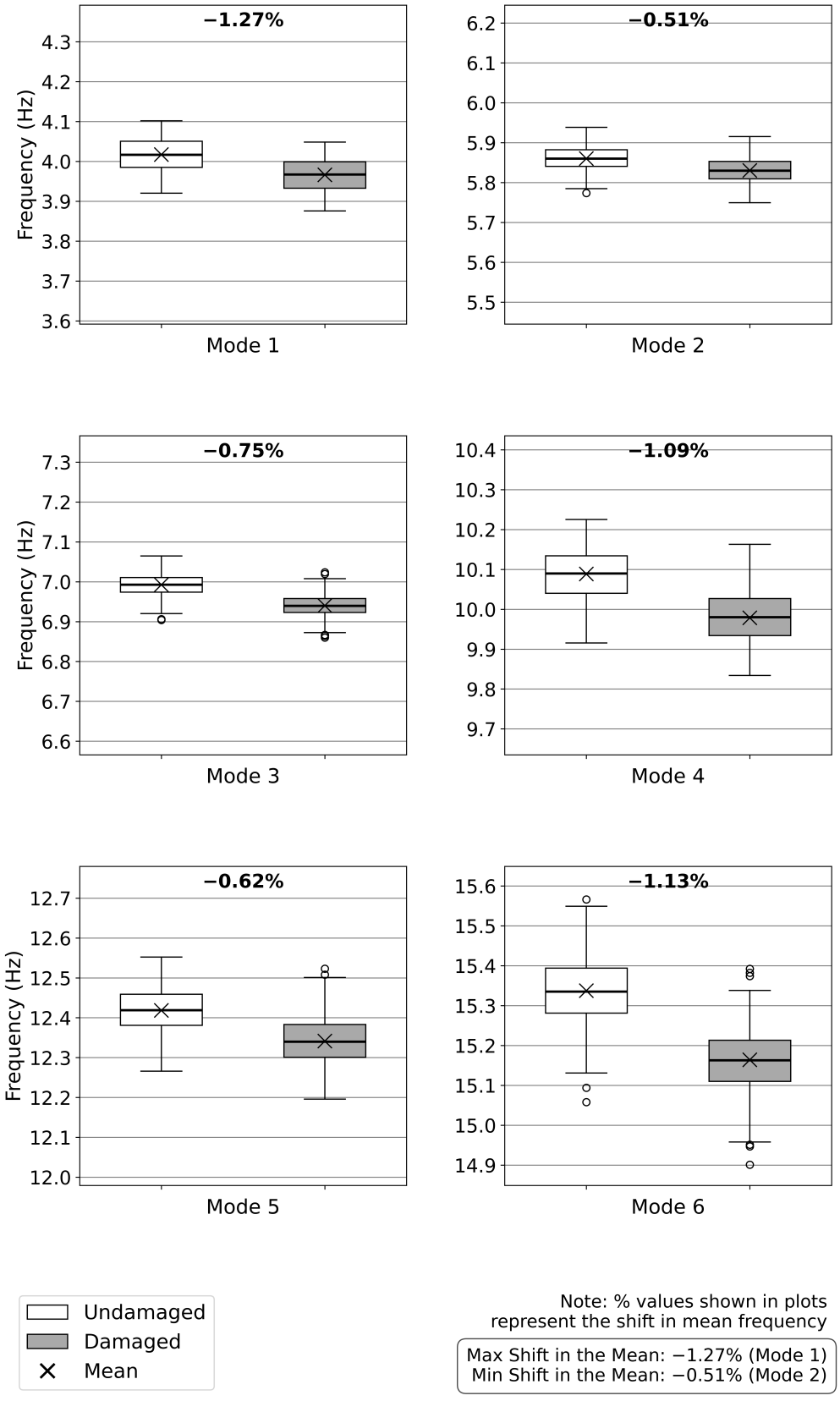

Figure 7. Box plots of natural frequency distributions for the undamaged and 40% pier stiffness reduction cases, both under thermal loads.

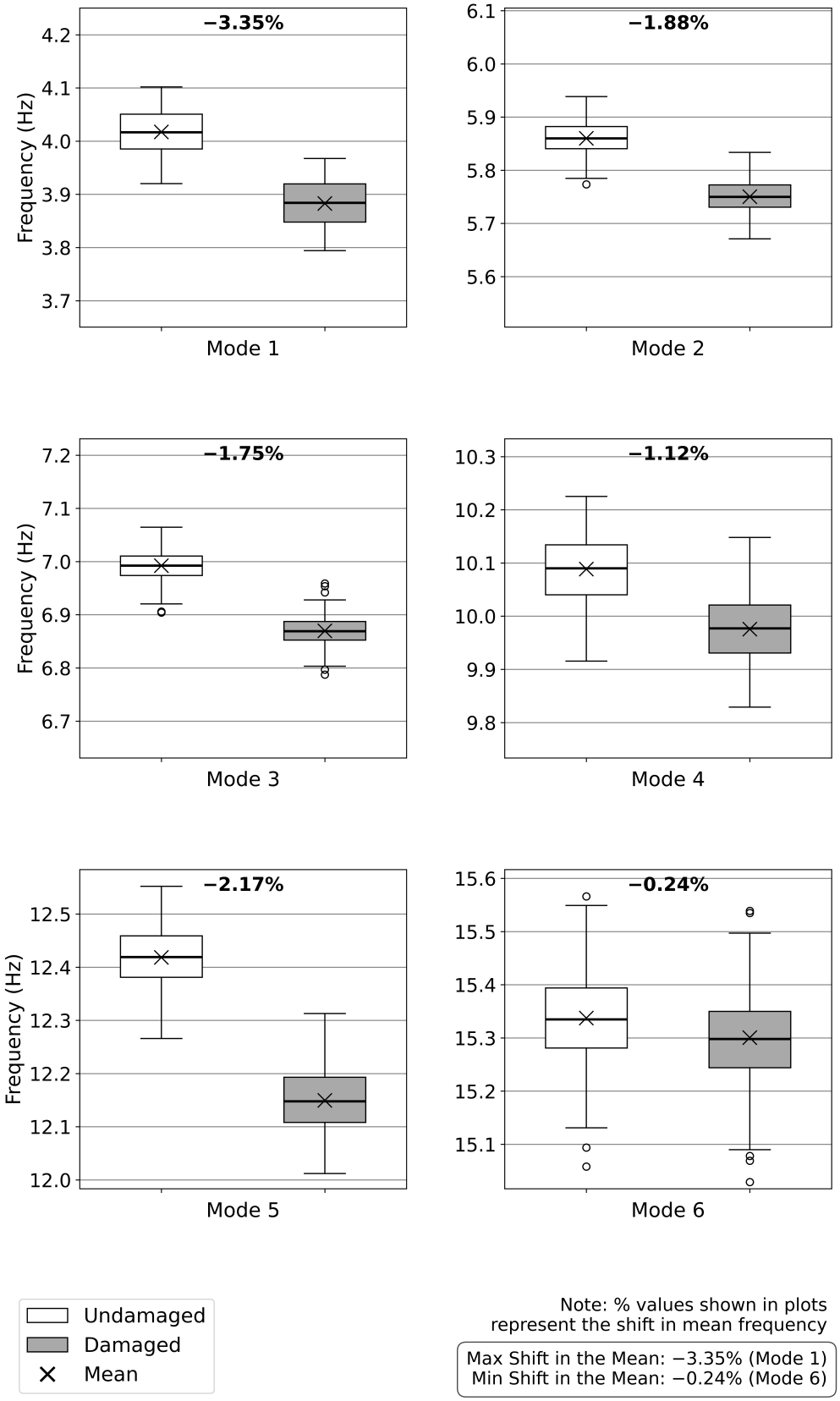

Figure 8. Box plots of natural frequency distributions for the undamaged and 40% mid-deck stiffness reduction cases, both under thermal loads.

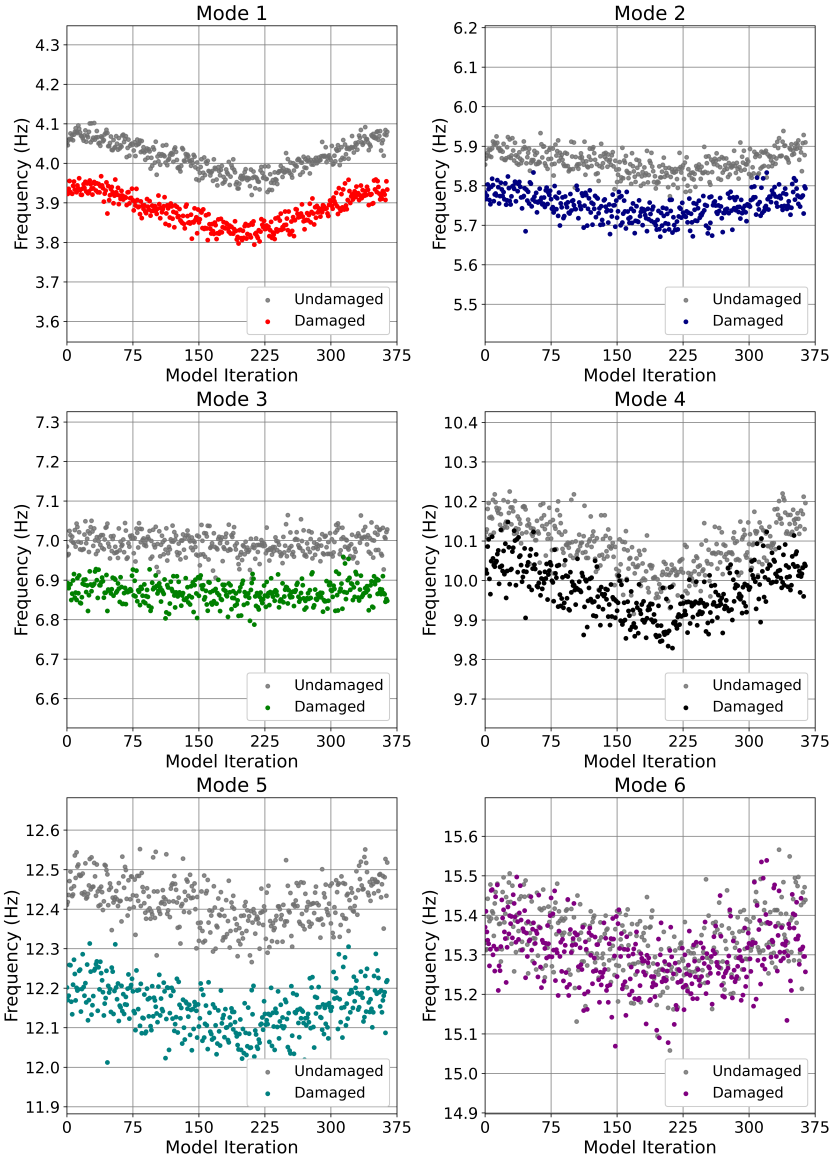

Figure 9. Scatter diagrams of natural frequency distributions for the undamaged and 40% mid-deck stiffness reduction cases, both under thermal loads.

3.2.3. Pier Settlement (Case 3)

In contrast to the stiffness reduction cases, simulated pier settlement yielded minimal frequency variations across all modes and severities. Even at the maximum settlement of 10 cm, mean frequency shifts ranged from −0.06% to −0.71% across the six modes, with the largest shift observed in Mode 1 and the smallest in Mode 3 (Figure 10), with all remaining modes exhibiting minor change. To contextualise these shifts, the comparison between the thermal and non-thermal results of this damage case demonstrated that temperature alone accounts for mean frequency reductions of −0.13% to −1.33% across the six modes. The settlement-induced shifts are therefore of comparable or lesser magnitude to those produced by seasonal temperature variation alone, meaning that the damage signal is largely indistinguishable from normal environmental fluctuation within the thermally loaded distributions. Consequently, the damaged and undamaged frequency distributions showed considerable overlap, as is evident in Figure 11.

Table 2 presents a direct comparison between the synthetic results for a 3 cm settlement and the empirical data recorded during the S101 destructive testing campaign [30]. The discrepancy is considerable; the physical experiment recorded frequency reductions of

between 9.23% and 23.09%, whereas the numerical simulation produced reductions of less than 1% across all mode shapes. This order-of-magnitude difference indicates that the linear FE modelling approach, which applies geometric displacement without inducing internal stress redistribution or progressive cracking, fails to replicate the non-linear damage mechanisms present in the real structure. This limitation is consistent with observations by Karimi and Gündel [31], who noted that simplified linear models cannot adequately capture crack opening or stress redistribution under settlement loading. Future research should therefore prioritise high-fidelity non-linear modelling for this damage class, as the current implementation produces synthetic data that is insufficiently discriminative for reliable algorithm training, as is further evidenced in Section 3.3.

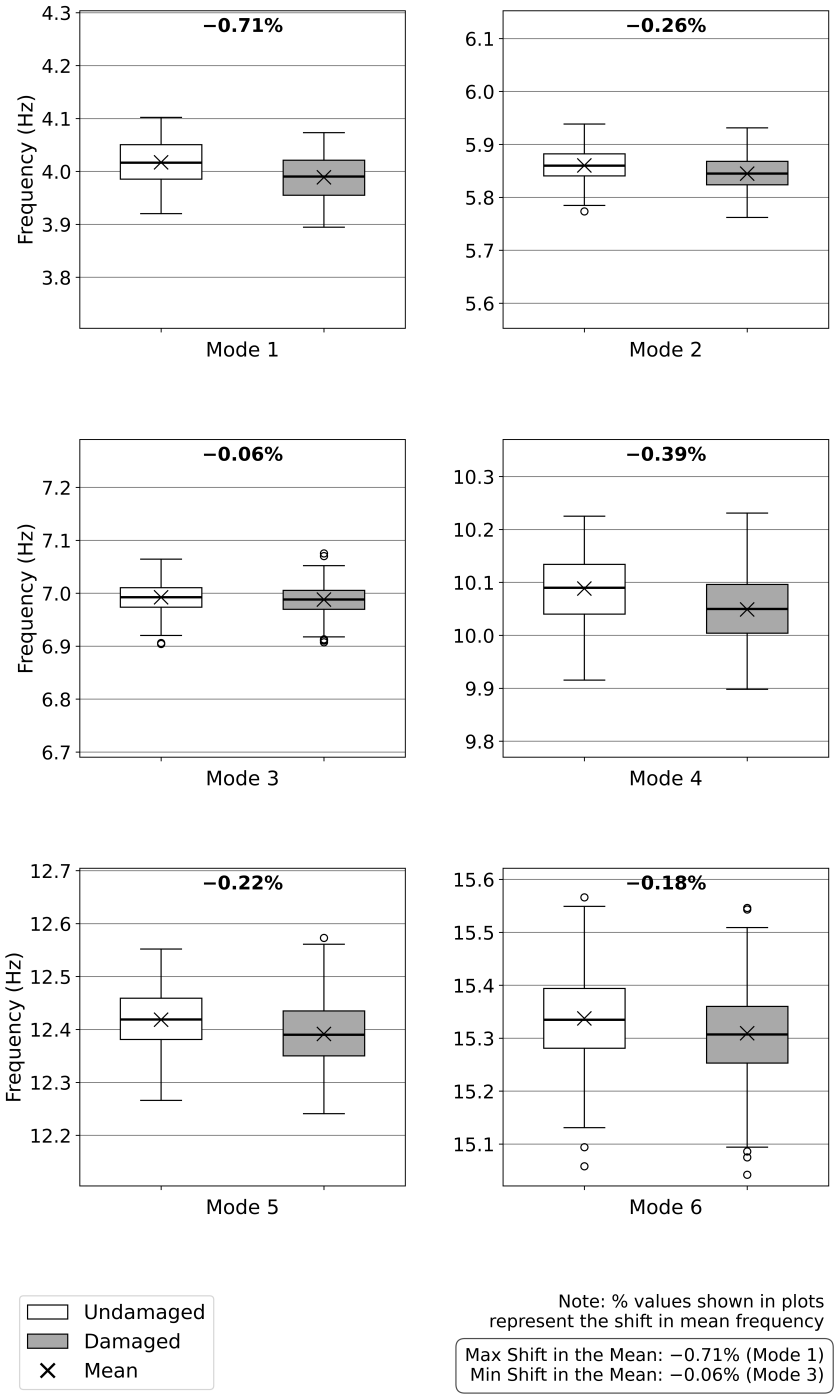

Figure 10. Box plots of natural frequency distributions for the undamaged and 10 cm pier settlement cases, both under thermal loads.

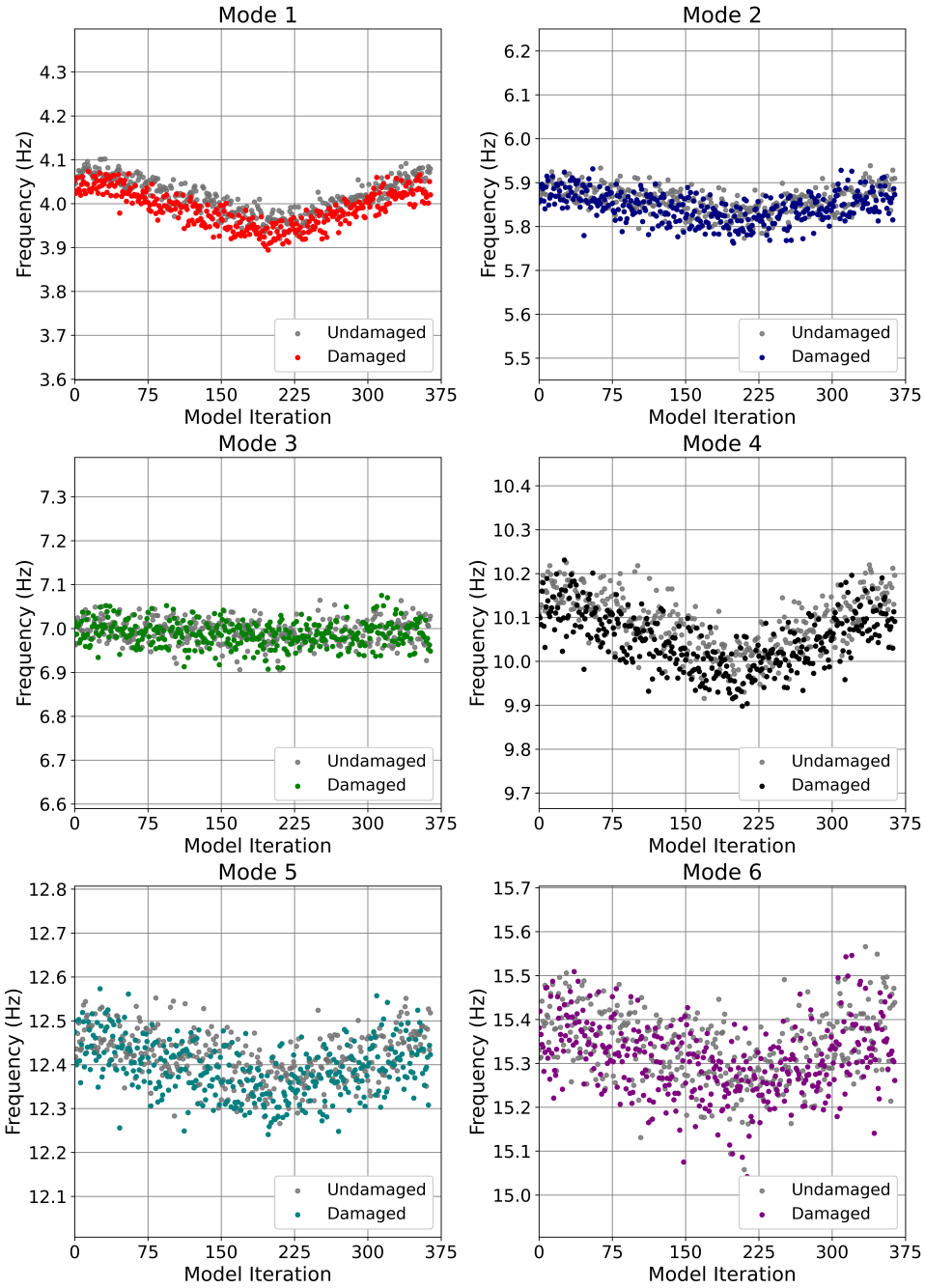

Figure 11. Scatter diagrams of natural frequency distributions for the undamaged and 10 cm pier settlement cases, both under thermal loads.

Table 2. Comparison of natural frequency percentage change between experimental data (2.7 cm settlement/'Damage 5') [30] and synthetic simulation (3 cm settlement).

Mode Shape	Experimental Change	Synthetic Change
1st Bending	−9.23%	−0.60%
1st Torsion	−17.31%	−0.48%
2nd Bending	−15.40%	−0.12%
2nd Torsion	−23.09%	−0.21%

3.2.4. Summary of Results

A numerical summary of the mean modal frequency shifts underlying the preceding discussion is provided in Table 3, covering all eighteen damage scenarios and the six modes considered throughout this study.

Table 3. Mean modal frequency shift (%) induced by structural damage, relative to the undamaged baseline, across all damage cases, severities and thermal conditions.

Damage Case	Thermal *	Sev.	M1	M2	M3	M4	M5	M6
Pier Stiffness Reduction	N	40%	−1.14	−0.47	−0.73	−1.02	−0.58	−1.10
	N	20%	−0.51	−0.21	−0.35	−0.46	−0.26	−0.44
	N	10%	−0.25	−0.10	−0.17	−0.22	−0.13	−0.20
	Y	40%	−1.27	−0.51	−0.75	−1.09	−0.62	−1.13
	Y	20%	−0.63	−0.25	−0.36	−0.52	−0.29	−0.47
	Y	10%	−0.36	−0.15	−0.19	−0.28	−0.16	−0.23
Mid Deck Stiffness Reduction	N	40%	−3.15	−1.82	−1.74	−1.06	−2.12	−0.21
	N	20%	−1.29	−0.75	−0.69	−0.42	−0.90	−0.08
	N	10%	−0.59	−0.35	−0.31	−0.19	−0.42	−0.04
	Y	40%	−3.35	−1.88	−1.75	−1.12	−2.17	−0.24
	Y	20%	−1.43	−0.80	−0.70	−0.48	−0.94	−0.11
	Y	10%	−0.71	−0.39	−0.33	−0.25	−0.46	−0.07
Pier Settlement	N	10 cm	−0.58	−0.22	−0.05	−0.32	−0.18	−0.15
	N	5 cm	−0.29	−0.11	−0.02	−0.16	−0.09	−0.07
	N	3 cm	−0.18	−0.07	−0.01	−0.10	−0.06	−0.05
	Y	10 cm	−0.71	−0.26	−0.06	−0.39	−0.22	−0.18
	Y	5 cm	−0.41	−0.16	−0.03	−0.23	−0.13	−0.11
	Y	3 cm	−0.29	−0.11	−0.02	−0.16	−0.10	−0.08

* Y = with thermal loads applied; N = non-thermal. M1–M6 denote Modes 1–6.

3.2.5. Validation of Synthetic Frequency Variability

As an independent check on the statistical character of the synthetic variability, the distributions of the six natural frequencies were tested for normality across all eighteen damage scenarios, for both the damaged and undamaged samples within each scenario, a property commonly assumed for identified natural frequencies in automated modal identification frameworks applied to instrumented bridges [44]. Since the undamaged baseline dataset is shared across scenarios within each thermal condition, this yielded 120 statistically independent distributions (60 non-thermal, 60 thermal). Under non-thermal conditions, no distribution departed from normality in any mode, case or severity (Shapiro–Wilk $p > 0.05$ throughout). Under thermal loading, Mode 1 departed from normality in every case tested (10 of 10 independent distributions), and Mode 4 in 4 of 10 cases, confined specifically to the pier stiffness reduction scenarios; the remaining four modes showed no departure from normality under any condition. This pattern is consistent with physically meaningful behaviour: Mode 1 and, for the pier case specifically, Mode 4 were identified in Section 3.2 as the modes most sensitive to thermal loading, and the shape of the empirical temperature record used to derive thermal loads is not itself required to be normal. The observed departure from normality is therefore plausibly attributable to this asymmetry propagating specifically into the modes most sensitive to temperature, rather than to an arbitrary artefact of the generation process.

3.3. Machine Learning Classification Performance

The classification results for both algorithms across all damage cases and thermal conditions are summarised in Table 4. The supervised NN consistently outperformed the unsupervised k-means algorithm across all scenarios, with the advantage being most pronounced under thermal loading and at lower damage severities. The supervised approach benefited directly from the availability of labelled synthetic data, allowing it to learn non-linear relationships between frequency shifts and damage states [9]. In contrast, the unsupervised k-means algorithm relied on the geometric separation of clusters in frequency space, which became increasingly indistinct as thermal variance caused distributions to overlap.

Table 4. Accuracy and AUC (supervised NN) and RI (unsupervised k-means) results (mean ± standard deviation across ten runs) across damage cases with and without thermal loads.

Damage Case	Thermal *	Sev.	NN		K-Means
			Accuracy	AUC	RI
(1) Pier Stiffness Reduction	N	40%	1.00 ± 0.00	1.00 ± 0.00	0.83 ± 0.03
	N	20%	0.95 ± 0.02	0.99 ± 0.01	0.58 ± 0.02
	N	10%	0.76 ± 0.07	0.83 ± 0.07	0.52 ± 0.01
	Y	40%	0.92 ± 0.05	0.97 ± 0.02	0.74 ± 0.03
	Y	20%	0.71 ± 0.03	0.79 ± 0.03	0.56 ± 0.02
	Y	10%	0.60 ± 0.02	0.66 ± 0.03	0.52 ± 0.01
(2) Mid Deck Stiffness Reduction	N	40%	1.00 ± 0.00	1.00 ± 0.00	0.99 ± 0.01
	N	20%	0.99 ± 0.04	1.00 ± 0.01	0.66 ± 0.02
	N	10%	0.92 ± 0.03	0.97 ± 0.01	0.53 ± 0.01
	Y	40%	1.00 ± 0.00	1.00 ± 0.00	0.93 ± 0.03
	Y	20%	0.99 ± 0.01	1.00 ± 0.00	0.61 ± 0.02
	Y	10%	0.90 ± 0.04	0.97 ± 0.02	0.53 ± 0.01
(3) Pier Settlement	N	10 cm	0.94 ± 0.01	0.99 ± 0.00	0.53 ± 0.01
	N	5 cm	0.80 ± 0.03	0.88 ± 0.02	0.51 ± 0.01
	N	3 cm	0.66 ± 0.06	0.73 ± 0.06	0.50 ± 0.00
	Y	10 cm	0.62 ± 0.02	0.71 ± 0.02	0.53 ± 0.01
	Y	5 cm	0.58 ± 0.02	0.63 ± 0.02	0.51 ± 0.01
	Y	3 cm	0.55 ± 0.01	0.57 ± 0.03	0.51 ± 0.01

* Y = with thermal loads applied; N = non-thermal. All values are the mean ± standard deviation across ten independent runs with different random train/test partitions.

For mid-deck stiffness reduction (Case 2), the NN achieved accuracies of 1.00 in both thermal and non-thermal conditions at 40% severity, reflecting the large and well-separated frequency shifts produced by this damage case. Performance remained strong even at 10% severity (0.92 under non-thermal conditions and 0.90 under thermal loading), suggesting that deck stiffness loss is reliably detectable from modal frequency data alone. For pier stiffness reduction (Case 1), classification accuracy degraded more significantly with decreasing severity and the introduction of thermal loads. At 10% severity, accuracy fell from 0.76 under non-thermal conditions to 0.60 under thermal loading, approaching the threshold of reliable detection.

The settlement case (Case 3) proved effectively undetectable by both algorithms under thermal loading. NN accuracies of 0.55 to 0.62 and Adjusted Rand Index values at 0.06 or lower confirm performance close to that of random assignment, a direct consequence of the negligible frequency shifts produced by the linear settlement model being fully masked by thermal variability. Even under non-thermal conditions, where the NN achieved

0.94 accuracy at 10 cm severity, performance at 3 cm settlement fell to 0.66, indicating that the synthetic data for this case lacks the discriminative signal necessary for reliable damage identification.

Extended classification metrics for the supervised NN, including precision, recall, F1-score and balanced accuracy averaged across ten repeated runs for each scenario, are reported in Appendix A. These broadly mirror the accuracy-based results discussed above, with all four metrics exceeding 0.98 for the well-separated scenarios and declining in step with accuracy as severity decreased or thermal loading was introduced. In the weaker-performing scenarios, recall consistently exceeded precision, for example 0.64 against 0.60 for pier stiffness reduction at 10% severity under thermal loading. This suggests the classifier's errors leaned toward false positives rather than false negatives, a bias that is preferable in a structural safety context.

Convergence and overfitting behaviour were also assessed across the repeated runs. The optimiser converged within the 1000-iteration limit in 179 of the 180 total runs. The sole exception occurred in the weakest-performing scenario (pier settlement at 10 cm under thermal loading). The gap between training and testing accuracy remained small across all scenarios, ranging from -3.7 to $+2.5$ percentage points, indicating no systematic overfitting.

Equivalent clustering diagnostics for k-means, including the Adjusted Rand Index, Normalised Mutual Information and silhouette score, are reported alongside these in Appendix A. The Adjusted Rand Index consistently falls below the Rand Index reported in Table 4, for example from 0.83 to 0.65 for pier stiffness reduction at 40% severity under non-thermal conditions, and collapses to 0.06 or lower in every scenario where the Rand Index sits close to 0.50. This confirms that the higher Rand Index values in weaker scenarios reflect the baseline inflation inherent to the metric rather than genuine cluster separation.

3.3.1. Gaussian Mixture Model

To provide a more rigorous benchmark for the unsupervised comparison, a Gaussian Mixture Model (GMM) with full covariance was also evaluated using the same repeated-run protocol, with results reported in Appendix A. In the well-separated scenarios, GMM substantially outperformed k-means, for example achieving a perfect Adjusted Rand Index for mid-deck stiffness reduction at 20% severity under non-thermal conditions compared to 0.33 for k-means, consistent with its ability to model non-spherical cluster shapes. This advantage did not extend to the more challenging scenarios: in several thermally-affected cases GMM performed no better than, or slightly worse than, k-means, for example returning an Adjusted Rand Index of -0.00 for mid-deck stiffness reduction at 10% severity under thermal loading, indistinguishable from random assignment. One scenario (mid-deck stiffness reduction at 20% severity under thermal loading) showed pronounced run-to-run instability, with an Adjusted Rand Index of 0.53 against a standard deviation of 0.48 across the ten repeated runs, indicating that GMM's additional flexibility can produce notably different local optima under noisy conditions rather than more reliable performance. Both unsupervised methods degrade sharply under thermal masking and low damage severity, reinforcing that the supervised NN's advantage in these conditions is not an artefact of an unrepresentative baseline.

3.3.2. Learning Curve Convergence

The learning curve for Case 1 at 40% severity under thermal loads is illustrated in Figure 12, revealing that whilst overfitting diminished as the training set size increased towards 730 samples, the training and validation scores had not fully converged at this dataset size. This supports the argument by Izonin et al. [45] regarding the challenges of limited datasets and indicates that the current scale of synthetic generation, whilst sufficient

to demonstrate the framework's utility, would benefit from expansion to achieve a more stable and generalisable classifier. To quantify this, an inverse power-law model was fitted to the cross-validated learning curve. A sensitivity analysis across plausible asymptotic performance levels (0.95–1.00) indicated that convergence would require approximately 2400 to 3650 samples, roughly three to five times the current dataset size, a substantial but achievable computational undertaking. This estimate is indicative rather than precise, given the limited data available for the extrapolation. Data augmentation techniques or significantly larger synthetic generation campaigns are likely necessary to realise the full potential of the proposed framework.

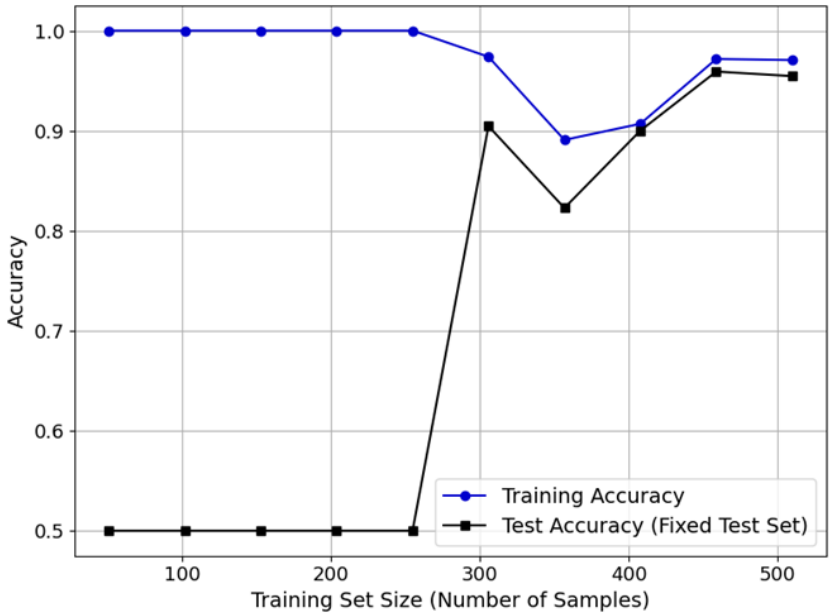

Figure 12. Neural network learning curve for Damage Case 1 (40% severity, thermal load).

3.4. Modelling Limitations

Whilst the framework demonstrated clear utility for stiffness-based damage cases, three limitations warrant consideration. First, the linear approximation of pier settlement produced synthetic data that failed to reflect the physical reality of the damage mechanism, as evidenced by the order-of-magnitude discrepancy with experimental observations. This finding indicates that, although the proposed framework is capable of generating physically meaningful synthetic datasets for stiffness-related deterioration mechanisms, the current implementation is not suitable for damage processes dominated by highly non-linear structural behaviour. In particular, progressive pier settlement induces cracking, tendon force redistribution and changes in structural continuity that cannot be adequately reproduced through linear elastic finite element analysis. Consequently, the settlement scenarios included in the present dataset should be regarded primarily as illustrative examples demonstrating the limitations of the adopted modelling assumptions rather than representative benchmark cases for machine learning development. More broadly, this highlights that the applicability of synthetic datasets depends not only on the data generation framework itself but also on the capability of the underlying numerical model to faithfully reproduce the governing physical mechanisms of each damage type. Addressing this limitation requires the adoption of non-linear FE modelling techniques capable of simulating crack propagation, material degradation and internal stress redistribution, thereby enabling future synthetic datasets to represent a wider range of realistic bridge deterioration mechanisms.

Second, whilst this study utilised a steady-state uniform thermal profile, real structures experience complex spatial and temporal thermal gradients across their cross-sections. Incorporating non-uniform thermal distributions derived from more detailed thermomechanical analyses could further enhance the physical fidelity of the synthetic dataset and improve the realism of the masking effect captured in training data.

Finally, validation of the trained anomaly detection algorithms using the synthetic databases produced by the implemented framework against field monitoring data from similar operational structures remains essential to ensure that trained algorithms do not merely learn the intrinsic patterns of the synthetic generation process itself [46]. A classifier that performs well on synthetic data but fails to generalise to real sensor measurements would offer limited practical value for operational SHM. Cross-validation against in situ measurements from instrumented bridges would provide the necessary evidence that the learned frequency-damage relationships reflect genuine structural behaviour rather than artefacts of the modelling assumptions.

It should be noted that the adopted coefficient of variation (2%) was selected as a representative value to emulate the small measurement variability typically observed in vibration-based structural health monitoring systems, rather than as a calibrated physical parameter describing the uncertainty of the bridge itself. Consequently, the primary objective of the present work is not to identify an optimal uncertainty level, but rather to demonstrate a practical and reproducible framework for generating synthetic datasets under representative operational variability. A systematic sensitivity analysis considering alternative coefficients of variation would undoubtedly provide additional insight into the influence of measurement uncertainty on modal frequencies and machine learning performance. However, such an investigation would constitute a separate uncertainty quantification study and falls beyond the scope of the present work. Importantly, the proposed framework is fully parameterised, and the FEM model used can be provided upon request by the corresponding author, allowing users to readily regenerate or augment the published dataset using alternative uncertainty levels or stochastic models that better reflect the characteristics of their own monitoring systems or research objectives. In this sense, the published dataset should be regarded as a representative benchmark implementation of the proposed framework rather than as the only possible realisation.

4. Conclusions

This study has presented the first complete implementation of a previously proposed physics-based synthetic data generation framework for bridge SHM using a calibrated FE model of the S101 bridge. Building upon the conceptual methodology introduced in earlier work [16,17], the framework has been extended to incorporate stochastic aleatory variability, long-term thermal loading derived from a 20-year climate record and representative structural damage scenarios, resulting in the generation of an openly accessible, FAIR-compliant benchmark dataset. By combining these elements within a unified workflow, the present study moves beyond the conceptual development of the framework and delivers a practical resource comprising data from 7300 FE simulations across 20 structural states, specifically intended to support the training, validation and benchmarking of machine learning algorithms for vibration-based damage detection under bridge operational conditions.

The principal finding of this work is that environmental variability cannot be treated as a secondary consideration in the design of synthetic training datasets. Temperature-induced frequency shifts of up to −1.33% routinely exceeded the modal perturbations attributable to minor structural deterioration, rendering certain damage states statistically indistinguishable from healthy operation without explicit environmental compensation. This has a direct consequence for algorithm development: supervised classifiers trained

on non-thermal data demonstrated inflated performance under controlled conditions but degraded markedly when thermal loads were introduced, confirming that environmental representativeness is a prerequisite for generalisable damage detection models.

Damage detectability was found to be strongly governed by the structural role of the affected component. Mid-deck stiffness loss produced frequency deviations sufficiently large to support reliable classification across all tested severities, even under thermal loading, owing to the substantial contribution of the mid-span region to the global bending and torsional response of the structure. Pier stiffness reductions were detectably smaller in their modal influence and more readily obscured by thermal variance. The supervised multi-layer perceptron consistently outperformed the unsupervised k-means algorithm across all scenarios, with the advantage most pronounced where thermal noise and low damage severity caused frequency distributions to overlap, conditions under which geometric cluster separation becomes unreliable.

The linear FE modelling approach adopted here was found to be insufficient for simulating pier settlement. The order-of-magnitude discrepancy between the synthetic frequency shifts and those recorded during the S101 destructive testing campaign indicates that geometric displacement within a linear elastic framework does not reproduce the crack propagation and stress redistribution that govern the physical response under this failure mode. The settlement case therefore produced synthetic data that is insufficiently discriminative to serve as reliable training material within the current linear modelling framework. Synthetic generation of settlement scenarios is unlikely to yield useful training data until non-linear FE modelling is adopted for this damage class.

Taken together, these findings suggest that physics-based synthetic data generation, when grounded in a well-calibrated model and augmented with realistic environmental variability, offers a credible route to addressing the labelled data scarcity that currently constrains the development and benchmarking of ML-based SHM algorithms. The FAIR-compliant dataset produced in this study [43], comprising 7300 model iterations across 20 distinct structural states, is intended to serve as a standardised reference against which future classifiers can be evaluated under conditions of combined structural and environmental uncertainty.

Future work should prioritise the adoption of non-linear FE modelling to improve the physical fidelity of settlement and cracking scenarios, incorporating material behaviours such as progressive crack opening, corrosion-induced voiding and tendon yielding in place of the simplified stiffness reductions employed within this study. A further priority is the transition from binary classification towards algorithms capable of localising damage and quantifying severity. A broader range of unsupervised and anomaly-detection benchmarks, including Isolation Forests, One-Class SVMs and autoencoder-based approaches, would further contextualise the performance of the classifiers presented in this study. Future work should also investigate the influence of different uncertainty models on the generated synthetic datasets by performing systematic sensitivity analyses with varying coefficients of variation and alternative uncertainty representations. A more rigorous treatment distinguishing measurement uncertainty, parameter uncertainty and model-form uncertainty could further improve the physical realism of the synthetic data and quantify the robustness of machine learning algorithms to different sources of variability, thereby extending the proposed framework towards a comprehensive uncertainty quantification methodology.

Thermal modelling should similarly advance beyond uniform steady-state profiles towards spatially non-uniform temperature distributions derived from empirical measurements or high-fidelity thermomechanical simulations, capturing the two-dimensional and three-dimensional temperature gradients that develop across bridge decks, webs and piers under realistic environmental conditions. Such temperature fields would enable a more

faithful representation of the differential thermal strains and stress redistributions that influence the dynamic behaviour of bridge structures. Furthermore, future research should investigate the integration of measured temperature data with machine learning algorithms to compensate for environmental effects during damage identification, following recent developments in temperature-aware structural health monitoring [47].

Ultimately, the implemented synthetic data generation framework has the potential to address the labelled data scarcity that currently constrains the deployment of ML-based SHM in practice. Although the dataset presented in this study is specific to the S101 bridge and cannot be directly generalised to other structures, the underlying methodology is readily transferable to other bridge typologies provided that a suitably calibrated finite element model is available. By combining structural calibration, stochastic uncertainty quantification, environmental loading, representative damage simulation and automated data generation, the framework can be used to produce bridge-specific synthetic datasets for training and benchmarking machine learning algorithms. This generalisable workflow, originally proposed in previous work [16,17] and further developed in the present study, enables the systematic creation of digital twin "what-if" scenarios for a wide range of bridge structures. Such an approach could facilitate the development of algorithms capable of detecting and localising damage, thereby supporting more proactive and evidence-based maintenance of transport infrastructure.

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

Funding: This research received no external funding.

Data Availability Statement: The synthetic modal frequency dataset generated in this study, comprising 7300 model iterations and 43,800 individual frequency datapoints across 20 structural states, is publicly available [43] in accordance with FAIR principles [33].

Acknowledgments: The authors gratefully acknowledge the contribution of David Stacey for his insight during the development of this research, particularly in relation to the sourcing of the historical temperature data. In preparing this manuscript, the authors employed generative AI tools (ChatGPT version GPT-5.6 Luna) to support the structuring and linguistic polishing of text. Following the use of these tools, the authors carefully reviewed and revised the material as necessary and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

ARI	Adjusted Rand Index
AUC	Area Under the Curve
FAIR	Findable, Accessible, Interoperable, Reusable
FE	Finite Element
GMM	Gaussian Mixture Model
ML	Machine Learning
NMI	Normalised Mutual Information
NN	Neural Network
PCA	Principal Component Analysis
RI	Rand Index
ROC	Receiver Operating Characteristic
SHM	Structural Health Monitoring

Appendix A. Extended Classification Performance Metrics

Table A1. NN hyperparameter tuning & architecture selection: Best cross-validated architecture and regularisation strength versus the adopted configuration (100, 50, 30, 15; $\alpha = 1 \times 10^{-3}$), five-fold stratified cross-validation, balanced accuracy.

Damage Case	Thermal *	Sev.	Best Arch.	Best α	Best CV	Adopted CV
(1) Pier Stiffness Reduction	N	40%	(50)	1×10^{-3}	1.00	0.97 ± 0.05
	N	20%	(100, 50, 30, 15)	1×10^{-4}	0.95	0.94 ± 0.02
	N	10%	(100, 50, 30, 15)	1×10^{-2}	0.77	0.77 ± 0.04
	Y	40%	(50)	1×10^{-4}	0.94	0.92 ± 0.05
	Y	20%	(50)	1×10^{-3}	0.70	0.68 ± 0.04
	Y	10%	(100, 50)	1×10^{-3}	0.60	0.56 ± 0.02
(2) Mid Deck Stiffness Reduction	N	40%	(50)	1×10^{-2}	1.00	1.00 ± 0.00
	N	20%	(50)	1×10^{-4}	1.00	0.99 ± 0.01
	N	10%	(150, 75, 40, 20)	1×10^{-4}	0.93	0.89 ± 0.07
	Y	40%	(100, 50)	1×10^{-3}	1.00	1.00 ± 0.00
	Y	20%	(50)	1×10^{-3}	1.00	1.00 ± 0.00
	Y	10%	(100, 50)	1×10^{-3}	0.93	0.85 ± 0.08
(3) Pier Settlement	N	10 cm	(100,50)	1×10^{-2}	0.95	0.89 ± 0.08
	N	5 cm	(50)	1×10^{-2}	0.81	0.78 ± 0.04
	N	3 cm	(50)	1×10^{-4}	0.68	0.65 ± 0.03
	Y	10 cm	(100, 50, 30, 15)	1×10^{-4}	0.61	0.59 ± 0.05
	Y	5 cm	(150, 75, 40, 20)	1×10^{-3}	0.60	0.55 ± 0.05
	Y	3 cm	(50)	1×10^{-2}	0.53	0.52 ± 0.06

* Y = with thermal loads applied; N = non-thermal. Best CV and Adopted CV are five-fold cross-validated balanced accuracy (mean ± standard deviation for the adopted architecture).

Table A2. Supervised NN precision, recall, F1-score and balanced accuracy (mean ± standard deviation, ten runs) across damage cases with and without thermal loads.

Damage Case	Thermal *	Sev.	Precision	Recall	F1-Score	Bal. Acc.
(1) Pier Stiffness Reduction	N	40%	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	N	20%	0.95 ± 0.02	0.95 ± 0.03	0.95 ± 0.02	0.95 ± 0.02
	N	10%	0.75 ± 0.07	0.77 ± 0.08	0.76 ± 0.07	0.76 ± 0.07
	Y	40%	0.92 ± 0.05	0.92 ± 0.05	0.92 ± 0.05	0.92 ± 0.05
	Y	20%	0.71 ± 0.05	0.72 ± 0.05	0.71 ± 0.03	0.71 ± 0.03
	Y	10%	0.60 ± 0.03	0.64 ± 0.07	0.61 ± 0.03	0.60 ± 0.02
(2) Mid Deck Stiffness Reduction	N	40%	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	N	20%	0.99 ± 0.04	0.99 ± 0.04	0.99 ± 0.04	0.99 ± 0.04
	N	10%	0.90 ± 0.03	0.93 ± 0.03	0.92 ± 0.03	0.92 ± 0.03
	Y	40%	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
	Y	20%	0.99 ± 0.02	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01
	Y	10%	0.91 ± 0.02	0.90 ± 0.06	0.90 ± 0.04	0.90 ± 0.04
(3) Pier Settlement	N	10 cm	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01	0.94 ± 0.01
	N	5 cm	0.80 ± 0.05	0.80 ± 0.03	0.80 ± 0.03	0.80 ± 0.03
	N	3 cm	0.67 ± 0.07	0.69 ± 0.11	0.67 ± 0.02	0.66 ± 0.06
	Y	10 cm	0.62 ± 0.03	0.64 ± 0.05	0.63 ± 0.02	0.62 ± 0.02
	Y	5 cm	0.58 ± 0.02	0.62 ± 0.12	0.59 ± 0.04	0.58 ± 0.02
	Y	3 cm	0.55 ± 0.02	0.59 ± 0.13	0.56 ± 0.06	0.55 ± 0.01

* Y = with thermal loads applied; N = non-thermal.

Table A3. Unsupervised k-means Adjusted Rand Index, Normalised Mutual Information and silhouette score (mean ± standard deviation, ten runs) across damage cases with and without thermal loads.

Damage Case	Thermal *	Sev.	ARI	NMI	Silhouette
(1) Pier Stiffness Reduction	N	40%	0.65 ± 0.07	0.56 ± 0.07	0.51 ± 0.01
	N	20%	0.16 ± 0.04	0.12 ± 0.03	0.41 ± 0.01
	N	10%	0.05 ± 0.02	0.04 ± 0.01	0.40 ± 0.01
	Y	40%	0.47 ± 0.07	0.38 ± 0.06	0.48 ± 0.02
	Y	20%	0.11 ± 0.03	0.09 ± 0.02	0.43 ± 0.02
	Y	10%	0.05 ± 0.02	0.04 ± 0.02	0.42 ± 0.01
(2) Mid Deck Stiffness Reduction	N	40%	0.97 ± 0.02	0.95 ± 0.04	0.65 ± 0.01
	N	20%	0.33 ± 0.05	0.26 ± 0.04	0.39 ± 0.02
	N	10%	0.06 ± 0.02	0.05 ± 0.02	0.39 ± 0.01
	Y	40%	0.85 ± 0.06	0.77 ± 0.07	0.57 ± 0.01
	Y	20%	0.22 ± 0.04	0.17 ± 0.03	0.42 ± 0.02
	Y	10%	0.06 ± 0.02	0.05 ± 0.02	0.42 ± 0.01
(3) Pier Settlement	N	10 cm	0.05 ± 0.02	0.04 ± 0.01	0.39 ± 0.01
	N	5 cm	0.02 ± 0.01	0.01 ± 0.01	0.39 ± 0.01
	N	3 cm	0.01 ± 0.01	0.01 ± 0.01	0.39 ± 0.01
	Y	10 cm	0.06 ± 0.02	0.05 ± 0.01	0.42 ± 0.01
	Y	5 cm	0.03 ± 0.02	0.02 ± 0.01	0.42 ± 0.02
	Y	3 cm	0.02 ± 0.02	0.01 ± 0.01	0.42 ± 0.02

* Y = with thermal loads applied; N = non-thermal.

Table A4. Gaussian Mixture Model Adjusted Rand Index, Normalised Mutual Information and silhouette score (mean ± standard deviation, ten runs) across damage cases with and without thermal loads.

Damage Case	Thermal *	Sev.	ARI	NMI	Silhouette
(1) Pier Stiffness Reduction	N	40%	1.00 ± 0.00	1.00 ± 0.00	0.44 ± 0.03
	N	20%	0.74 ± 0.22	0.67 ± 0.22	0.18 ± 0.09
	N	10%	0.05 ± 0.03	0.04 ± 0.02	0.35 ± 0.02
	Y	40%	0.97 ± 0.01	0.95 ± 0.02	0.34 ± 0.02
	Y	20%	0.03 ± 0.01	0.03 ± 0.01	0.22 ± 0.02
	Y	10%	0.02 ± 0.01	0.02 ± 0.01	0.23 ± 0.02
(2) Mid Deck Stiffness Reduction	N	40%	1.00 ± 0.00	1.00 ± 0.00	0.65 ± 0.01
	N	20%	1.00 ± 0.00	1.00 ± 0.00	0.30 ± 0.03
	N	10%	0.09 ± 0.03	0.07 ± 0.03	0.36 ± 0.01
	Y	40%	1.00 ± 0.00	1.00 ± 0.00	0.56 ± 0.02
	Y	20%	0.53 ± 0.48	0.52 ± 0.48	0.22 ± 0.06
	Y	10%	−0.00 ± 0.00	0.00 ± 0.00	0.21 ± 0.02
(3) Pier Settlement	N	10 cm	0.06 ± 0.03	0.05 ± 0.02	0.36 ± 0.02
	N	5 cm	0.02 ± 0.02	0.02 ± 0.01	0.34 ± 0.02
	N	3 cm	0.01 ± 0.01	0.01 ± 0.01	0.34 ± 0.02
	Y	10 cm	0.06 ± 0.02	0.05 ± 0.01	0.24 ± 0.02
	Y	5 cm	0.02 ± 0.01	0.02 ± 0.01	0.23 ± 0.02
	Y	3 cm	0.01 ± 0.01	0.01 ± 0.01	0.23 ± 0.02

* Y = with thermal loads applied; N = non-thermal.

References

1. Kopiika, N.; Robery, P.; Ninic, J.; Mitoulis, S.A. Remaining Life of Ageing RC Infrastructure for Sustainable development—Deterioration Under Climate Change. *Case Stud. Constr. Mater.* **2025**, *22*, e04757. [CrossRef]

2. Shaban, N.; Arango, E.; Jiménez Rios, A.; Rincon, L.F.; Shang, Y.; Habeeb, B.; Martín-Pérez, B.; Ferreira, T.M.; Mendoza-Lugo, M.A.; Eudave, R.R.; et al. Perspectives on the resilience of existing bridges to short- and long-term impacts of climate change. *Struct. Infrastruct. Eng.* **2026**, 1–26. [CrossRef]
3. Dong, C.Z.; Catbas, F.N. A review of computer vision–based structural health monitoring at local and global levels. *Struct. Health Monit.* **2021**, *20*, 692–743. [CrossRef]
4. Housner, G.W.; Bergman, L.A.; Caughey, T.K.; Chassiakos, A.G.; Claus, R.O.; Masri, S.F.; Skelton, R.E.; Soong, T.T.; Spencer, B.F.; Yao, J.T.P. Structural Control: Past, Present, and Future. *J. Eng. Mech.* **1997**, *123*, 897–971. [CrossRef]
5. Farrar, C.R.; Worden, K. An introduction to structural health monitoring. *Phil. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2007**, *365*, 303–315. [CrossRef] [PubMed]
6. Sohn, H.; Farrar, C.R.; Hemez, F.M.; Czarnecki, J.J. A Review of Structural Health Monitoring Literature 1996–2001. In Proceedings of the Third World Conference on Structural Control. Los Alamos National Laboratory, Como, Italy, 7–11 April 2002.
7. Jiménez Rios, A.; Plevris, V.; Nogal, M. Bridge management through digital twin-based anomaly detection systems: A systematic review. *Front. Built Environ.* **2023**, *9*, 1176621. [CrossRef]
8. Jiménez Rios, A.; Ben Seghier, M.E.A.; Plevris, V.; Dai, J. Explainable ensemble learning framework for estimating corrosion rate in suspension bridge main cables. *Results Eng.* **2024**, *23*, 102723. [CrossRef]
9. Zhang, X.; Lu, Y. Robustness of wavelet features for damage detection with neural network and digital twin. *Proc. Inst. Civ. Eng.-Eng. Comput. Mech.* **2024**, *177*, 83–101. [CrossRef]
10. Le, N.; Keenan, M.; Nguyen, A.; Ghazvineh, S.; Yu, Y.; Li, J.; Manalo, A. A supervised machine learning approach for structural overload classification in railway bridges using weigh-in-motion data. *Structures* **2025**, *71*, 108005. [CrossRef]
11. Malekloo, A.; Ozer, E.; AlHamaydeh, M.; Girolami, M. Machine learning and structural health monitoring overview with emerging technology and high-dimensional data source highlights. *Struct. Health Monit.* **2022**, *21*, 1906–1955. [CrossRef]
12. Eltouny, K.; Gomaa, M.; Liang, X. Unsupervised Learning Methods for Data-Driven Vibration-Based Structural Health Monitoring: A Review. *Sensors* **2023**, *23*, 3290. [CrossRef] [PubMed]
13. Wang, Z.; Cha, Y. Unsupervised machine and deep learning methods for structural damage detection: A comparative study. *Eng. Rep.* **2022**, *7*, e12551. [CrossRef]
14. Huang, M.; Wan, N.; Zhu, H. Reconstruction of Structural Acceleration Response Based on CNN-BiGRU with Squeeze-and-Excitation Under Environmental Temperature Effects. *J. Civ. Struct. Health Monit.* **2025**, *15*, 985–1003. [CrossRef]
15. Wang, L.; Wang, X.; Zhang, J.; Wang, J.; Yu, H. A self-supervised learning-based approach for detection and classification of dam deformation monitoring abnormal data with imaging time series. *Structures* **2024**, *68*, 107148. [CrossRef]
16. Jiménez Rios, A.; Plevris, V.; Nogal, M. Synthetic Data Generation for the Creation of Bridge Digital Twins What-if Scenarios. In Proceedings of the ECCOMAS Thematic Conference on Computational Methods in Structural Dynamics and Earthquake Engineering, Athens, Greece, 12–14 June 2023. [CrossRef]
17. Jiménez Rios, A.; Plevris, V.; Nogal, M. Uncertainties in the synthetic data generation for the creation of bridge digital twins. In Proceedings of the 5th International Conference on Uncertainty Quantification in Computational Science and Engineering, Athens, Greece, 12–14 June 2023. [CrossRef]
18. Wang, X.; Gao, Q.; Liu, Y. Damage Detection of Bridges under Environmental Temperature Changes Using a Hybrid Method. *Sensors* **2020**, *20*, 3999. [CrossRef] [PubMed]
19. Kashyap, P.; Shivgan, K.; Patil, S.; Raja, B.R.; Mahajan, S.; Banerjee, S.; Tallur, S. Unsupervised deep learning framework for temperature-compensated damage assessment using ultrasonic guided waves on edge device. *Sci. Rep.* **2024**, *14*, 3751. [CrossRef] [PubMed]
20. Figueiredo, E.; Peres, N.; Moldovan, I.; Nasr, A. Impact of climate change on long-term damage detection for structural health monitoring of bridges. *Struct. Health Monit.* **2024**, *24*, 14759217231224254. [CrossRef]
21. Farrar, C.R.; Doebling, S.W.; Cornwell, P.J.; Straser, E.G. Variability of modal parameters measured on the Alamosa Canyon Bridge. In Proceedings of the XV International Modal Analysis Conference, Dearborn, MI, USA, 12–15 February 1996.
22. Xia, Y.; Chen, B.; Weng, S.; Ni, Y.Q.; Xu, Y.L. Temperature effect on vibration properties of civil structures: A literature review and case studies. *J. Civ. Struct. Health Monit.* **2012**, *2*, 29–46. [CrossRef]
23. Zolghadri, N.; Halling, M.; Barr, P.; Foust, N. *Effects of Temperature on Bridge Dynamic Properties*; Utah State University: Logan, UT, USA, 2015.
24. Wang, Z.; Liu, Y.; Liu, G.H.; Huang, J.Z.; Yao, X.J.; Gu, H.L.; Wang, B. Anomaly detection for bridge frequency abnormality via data neighborhood standardization-aided enhanced environment-frequency relationship. *Structures* **2026**, *83*, 110866. [CrossRef]
25. Peeters, B.; Maeck, J.; De Roeck, G. Vibration-based damage detection in civil engineering: Excitation sources and temperature effects. *Smart Mater. Struct.* **2001**, *10*, 518–527. [CrossRef]
26. Gong, X.; Song, X.; Li, G.; Xiong, W.; Cai, C. Deep learning based anomaly identification of temperature effects in bridge structural health monitoring data. *Structures* **2024**, *69*, 107478. [CrossRef]

27. Koo, K.Y.; Park, S.; Lee, J.J.; Yun, C.B. Automated Impedance-based Structural Health Monitoring Incorporating Effective Frequency Shift for Compensating Temperature Effects. *J. Intell. Mater. Syst. Struct.* **2009**, *20*, 367–377. [CrossRef]
28. Huang, M.; Lei, Y.; Li, X.; Gu, J. Damage Identification of Bridge Structures Considering Temperature Variations-Based SVM and MFO. *J. Aerosp. Eng.* **2021**, *34*, 04020113. [CrossRef]
29. Wu, Y.; Sicard, B.; Gadsden, S.A. A Review of Physics-Informed Machine Learning Methods with Applications to Condition Monitoring and Anomaly Detection. *arXiv* **2024**, arXiv:2401.11860. [CrossRef]
30. Siringoringo, D.M.; Fujino, Y.; Nagayama, T. Dynamic Characteristics of an Overpass Bridge in a Full-Scale Destructive Test. *J. Eng. Mech. ASCE* **2013**, *139*, 691–701. [CrossRef]
31. Karimi, S.; Gündel, M. Advancing Structural Health Monitoring in Civil Engineering with Greybox Modelling: A Review. *E-J. Nondestruct. Test.* **2024**, *29*, 1–9. [CrossRef] [PubMed]
32. Buckley, T.; Ghosh, B.; Pakrashi, V. A Feature Extraction & Selection Benchmark for Structural Health Monitoring. *Struct. Health Monit.* **2023**, *22*, 2082–2127. [CrossRef]
33. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]
34. *BS EN 18162:2026*; Building Information Modelling (BIM). Digital Twins Applied to the Built Environment. Concept and Definitions. British Standards Institution (BSI): London, UK, 2026.
35. Paolini, C.; Civera, M.; D'Angelo, M.; Giordano, P.F.; Borlenghi, P.; Ballio, F.; Chiaia, B.; Limongelli, M.P. Dataset of bridge collapses in italy spanning more than 25 years (2000–2025). *Data Brief* **2026**, *64*, 112375. [CrossRef] [PubMed]
36. Jiménez Rios, A.; O'Dwyer, D. External Post-tensioning System for the Strengthening of Historical Stone Masonry Bridges. In *Proceedings of the Structural Analysis of Historical Constructions*; Aguilar, R., Torrealva, D., Moreira, S., Pando, M.A., Ramos, L.F., Eds.; Springer: Cham, Switzerland, 2019; pp. 1566–1574.
37. Laflamme, S.; Blasch, E.; Ubertini, F.; Liu, Z.; Wertz, J.; Knott, C.; Cherry, M.; Lindgren, E.; Chang, F.K.; Kumar, A.; et al. Roadmap: Integrating artificial intelligence in structural health monitoring systems. *Meas. Sci. Technol.* **2026**, *37*, 103001. [CrossRef]
38. European Climate Assessment & Dataset. Air Temperature and Precipitation Series for the European Climate Assessment. (Station 16). ECA&D Project Portal, 2025. Available online: https://www.ecad.eu/download/millennium/millennium.php (accessed on 26 February 2025).
39. Döhler, M.; Hille, F.; Mevel, L.; Rücker, W. Structural health monitoring with statistical methods during progressive damage test of S101 Bridge. *Eng. Struct.* **2014**, *69*, 183–193. [CrossRef]
40. Jiménez Rios, A.; Demirlioglu, K.; Plevris, V.; Nogal, M. Towards the Creation of Synthetic Bridge Digital Twins What-if Scenarios: Study Case and Calibration. In Proceedings of the 9th ECCOMAS Congress, Lisboa, Portugal, 3–7 June 2024. [CrossRef]
41. Brambilla, M.; Chiariotti, P.; Cigada, A. On the Effect of Intra- and Inter-Node Sampling Variability on Operational Modal Parameters in a Digital MEMS-Based Accelerometer Sensor Network for SHM: A Preliminary Numerical Investigation. *Sensors* **2025**, *25*, 5044. [CrossRef] [PubMed]
42. Kamariotis, A.; Vlachas, K.; Ntertimanis, V.; Koune, I.; Cicirello, A.; Chatzi, E. On the Consistent Classification and Treatment of Uncertainties in Structural Health Monitoring Applications. *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng.* **2024**, *11*, 011108. [CrossRef]
43. Sinden, M.; Jiménez Rios, A. Dataset of the Synthetic Data Generation for the Prototyping of Bridge Damage Detection Algorithms. *Zenodo* **2026**. [CrossRef]
44. Cho, K.; Cho, J.R. Stochastic Subspace Identification-Based Automated Operational Modal Analysis Considering Modal Uncertainty. *Appl. Sci.* **2023**, *13*, 12274. [CrossRef]
45. Izonin, I.; Nesterenko, I.; Kazantzi, A.K.; Tkachenko, R.; Muzyka, R.; Mitoulis, S.A. Enhanced ANN-based ensemble method for bridge damage characterization using limited dataset. *Sci. Rep.* **2024**, *14*, 24395. [CrossRef] [PubMed]
46. Hao, S.; Han, W.; Jiang, T.; Li, Y.; Wu, H.; Zhong, C.; Zhou, Z.; Tang, H. Synthetic Data in AI: Challenges, Applications, and Ethical Implications. *arXiv* **2024**, arXiv:2401.01629. [CrossRef]
47. Huang, M.; Zhao, W.; Gu, J.; Lei, Y. Damage Identification of a Steel Frame Based on Integration of Time Series and Neural Network under Varying Temperatures. *Adv. Civ. Eng.* **2020**, *2020*, 4284381. [CrossRef]

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.