Data-Driven and Physics-Informed Neural Networks for Structural Health Monitoring of the Z24 Bridge

Abdellah Riyahi*, Mohammed Mestari, Bouchra Bouihi

2IACS Laboratory, ENSET Mohammedia, Hassan II University of Casablanca, Mohammedia, MOROCCO
*Corresponding author: a.riyahi@enset-media.ac.ma

ABSTRACT Structural Health Monitoring (SHM) is crucial for maintaining the sustainability and safety of civil infrastructure. The Z24 Bridge in Switzerland remains one of the benchmark datasets used to validate vibration-based damage detection methods. Traditional approaches based exclusively on modal parameters are frequently limited by data scarcity and environmental variability. Recent advances in artificial intelligence have enabled data-driven neural networks to learn discriminative features directly from raw measurements. Meanwhile, hybrid methods such as Physics-Informed Neural Networks (PINNs) incorporate governing physical laws into the learning process. This study presents a comparative analysis of three successive artificial Neural Network models (NN V1–V3) and One Physics-Informed Neural Network (PINN V1), all applied to the Z24 Bridge dataset. The NN models progressively improve in depth, optimization strategy, and regularization, achieving ≈97.7% validation accuracy and a macro AUC ≈1.00 with NN V3. However, they remain completely dependent on the quality and quantity of training data. In contrast, the PINN incorporates the differential equation of a damped oscillator into its loss function, balancing a data-driven term with a physics-based residual. This approach enables more stable learning with limited labeled data and ensures consistency with structural dynamics. Experimental results highlight the trade-off between accuracy and robustness: while NN V3 yields the highest predictive performance (≈97.7% validation accuracy, macro AUC ≈1.00), PINN V1 achieves slightly lower accuracy (≈92%) but offers improved stability and interpretability. This dual perspective demonstrates that hybrid physics-informed models provide a more reliable basis for decision-making in SHM. The findings underscore the potential of combining machine learning with physical knowledge, paving the way for future developments such as hybrid PINNs (HPINNs), multi-sensor integration, and high-performance computing deployment.

KEYWORDS Structural Health Monitoring; Z24 Bridge; Neural Networks; Physics-Informed Neural Networks; Vibration-Based Damage Detection; Hybrid Modeling.

1 INTRODUCTION

In modern civil engineering, Structural Health Monitoring (SHM) has emerged as a crucial challenge. Bridges and viaducts are major examples of aging infrastructure that are continuously subjected to traffic loads, environmental conditions, and material fatigue. This context has led to the development of advanced methods that enable early damage detection, as ensuring user safety and optimizing preventive maintenance are top priorities (Alla and Asadi, 2020). The Z24 Bridge in Switzerland remains one of the most iconic case studies and a major benchmark. This bridge generated a comprehensive dataset that has become an important reference for validating vibration-based SHM approaches, as it was instrumented and subjected to controlled damage scenarios prior to its demolition (Maeck and De Roeck, 2003).

Traditionally, SHM methods have relied on modal analysis techniques, including mode shapes, damping ratios, and natural frequency extraction. Although robust, these methods have limitations under real-world conditions due to measurement noise, environmen-

tal variability, and the difficulty of explicitly correlating modal changes with structural damage (Avci et al., 2021). Consequently, these limitations have paved the way for the rise of artificial intelligence (AI), particularly neural networks, which can automatically extract discriminative features from either preprocessed or raw signals (Plevris and Papazafeiropoulos, 2024).

When artificial neural networks are applied to SHM, they demonstrate significant potential for state classification, damage detection, and the prediction of nonlinear behaviors. Recent studies underline the urgency of reliable bridge monitoring and model credibility in civil infrastructure, reinforcing the relevance of this study (Seventekidis and Giagopoulos, 2021; Absor et al., 2024).

However, data-driven approaches depend entirely on the quality and availability of training datasets. Failure data are scarce in many cases, creating risks of overfitting and limited generalization capacity (Zhang et al., 2020). To overcome these limitations, hybrid ap-

proaches have emerged that integrate governing physical laws into the learning process, namely Physics-Informed Neural Networks (PINNs) (Raissi et al., 2019; Pawar et al., 2021; Oddiraju et al., 2025).

By combining the rigor of partial differential equations that describe structural dynamics with the predictive power of machine learning, PINNs represent a significant breakthrough. Unlike classical neural networks, whose loss functions depend solely on the error between predictions and observations, PINNs incorporate additional terms that enforce governing motion equations. This integration of physical constraints improves stability, reduces reliance on labeled data, and enhances interpretability (Karniadakis et al., 2021). This approach is especially promising in SHM because structural dynamic equations (velocities, displacements, and accelerations) are directly incorporated into the learning process (Absor et al., 2024).

Recent advances in vibration-based SHM combine fast data-driven classifiers with physics-guided learning. Compact time-series methods (e.g., convolutional or transformer-based approaches such as MiniROCKET) have demonstrated strong accuracy and efficiency on sensor streams, while PINNs have progressed from proof-of-concept studies to applications in structural dynamics, where enforcing governing equations improves interpretability and out-of-distribution robustness (Dempster et al., 2020; Karniadakis et al., 2021; Sun et al., 2022). In parallel, recent civil engineering studies emphasize the urgency of reliable bridge monitoring and highlight the need to evaluate not only predictive accuracy but also operational robustness and physical credibility (Purba et al., 2023; Absor et al., 2024).

Beyond single-degree-of-freedom (SDOF) settings, recent studies have extended PINNs to multi-degree-of-freedom (MDOF) structural dynamics, combining physics constraints with parametric or state inference based on multi-modal responses. These works report that physics-guided regularization can reduce overfitting and improve robustness under sparse labels, albeit at higher computational cost. Our study is complementary; we deliberately analyze the SDOF-regularized case on Z24 to isolate the benefits and limitations of a first-mode prior and to motivate future MDOF PINN developments (Felahi, 2023).

This study addresses that gap by:

- Conducting a systematic comparison between three lightweight data-driven MLP variants (NN V1–V3) and a physics-informed model (PINN V1) on the same Z24 dataset.
- Enforcing identical data handling and a fixed train/validation/test split across models to ensure fair comparability.

- Reporting multi-criteria metrics (accuracy, AUC, precision/recall, per-class confusion) together with physics diagnostics (e.g., residuals of the governing equation, power spectral density) and discussing the practical trade-offs between data-rich performance and physics-based interpretability.

Therefore, we study in a controlled and reproducible setting: (i) the progressive design choices in NN V1–V3 (depth, optimization, and regularization) and (ii) a PINN that explicitly integrates the motion law of a damped oscillator as a simplified prior for bridge dynamics, thereby coupling data-driven learning with physical constraints. The detailed quantitative findings and comparisons are presented in Sections 4–5.

The remainder of this paper is organized as follows: Section 1 reviews related work in SHM and PINNs; Section 2 describes the Z24 dataset; Section 3 details the developed models (NN and PINN) and the training methodology; Section 4 and 5 presents and discusses the experimental results; and Section 6 concludes with perspectives for future research.

We evaluate three purely data-driven classifiers—NN V1, NN V2, and NN V3—which represent successive improvements in depth, optimization, and regularization, alongside a physics-informed model (PINN V1) that embeds a damped-oscillator law consistent with bridge dynamics. The architectural and training details are provided in Section 3. A consolidated comparison of validation metrics and physics diagnostics is reported in Table 4 (Section 4), where we discuss accuracy, macro-F1, macro AUC, false negatives, and PINN residual/consistency indicators.

All models are trained and evaluated on windows extracted from the same vertical accelerometer (acc_09) to ensure like-for-like comparability across NN and PINN.

2 CASE STUDY: THE Z24 BRIDGE DATASET

One of the most studied experimental benchmarks in Structural Health Monitoring (SHM) is the Z24 Bridge in Switzerland. This post-tensioned concrete bridge, located near Koppigen on the Swiss highway A1, was 30 m long and 11.8 m wide, with three spans supported by two piers and abutments at each end. The bridge was extensively instrumented by the Swiss Federal Institute of Technology (ETH Zürich) and subjected to a sequence of controlled damage scenarios prior to its planned demolition in 1998, making it a unique large-scale testbed for SHM research (Maeck and De Roeck, 2003; Koenders and Papagiannakis, 2013).

During the monitoring campaign, which lasted over one year, the bridge was equipped with various sen-

sors, including temperature sensors, strain gauges, accelerometers, and displacement transducers. Measurements were collected under artificially induced damage states such as tendon rupture, pier settlement, and concrete cracking, as well as under varying operational and environmental conditions. This produced a wide range of structural responses, from healthy to progressively damaged states, making the dataset particularly valuable for evaluating machine learning–based damage detection algorithms (Avci et al., 2021; Civera and Surace, 2021).

The resulting dataset, distributed in MATLAB (.mat) format, is commonly referred to as the Z24 EMS dataset (Environmental and Modal Survey) for reproducibility. The EMS manifest ID used in this study is 4b8b4d63c00f3a81. It is organized into several configurations, each corresponding to a distinct test condition. Our research focuses on vibration response signals arranged into 17 different setups, each containing acceleration time series captured under intact or damaged structural states. The dataset includes both ambient vibration measurements and controlled excitations, enabling the evaluation of supervised and unsupervised learning frameworks (Avci et al., 2021; Eltouny et al., 2023).

The Z24 dataset's combination of progressive artificial damage introduction and long-term monitoring data is a key feature. This dual nature allows researchers to address two complementary SHM challenges: (i) determining the degree of damage in specific scenarios (e.g., pier settlement) and (ii) distinguishing natural fluctuations (e.g., due to temperature and humidity changes) from actual structural deterioration. Consequently, Z24 has been widely adopted as a reference dataset in studies using deep learning, Physics-Informed Neural Networks (PINNs), and time-series classification methods such as WaveNet and MiniRocket (Karniadakis et al., 2021; Dabbous et al., 2024).

One objective of this work is to exploit the vibration-based setups of the Z24 dataset to evaluate both conventional neural networks (NN V1–V3) and hybrid physics-informed approaches (PINN V1). Each setup is represented by multiple .mat files containing acceleration signals, with durations ranging from a few seconds to several minutes, depending on the testing configuration. A typical file structure includes metadata such as excitation type and structural state, along with multi-channel acceleration data sampled at high frequency.

The systematic organization of the dataset enables the creation of supervised classification tasks in which a structural state label is assigned to each input time series. To evaluate both classification accuracy and physical interpretability, we selected representative subsets that balance healthy and damaged conditions. As a re-
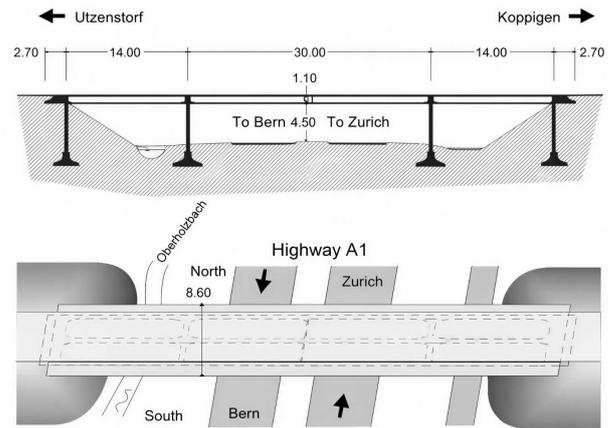

Figure 1. Front elevation (top) and plan view (bottom) of the Z24 Bridge. Redrawn based on structural descriptions from the SIMCES project at KU Leuven ((KU Leuven, 2024)).

sult, the Z24 dataset serves both as a historical benchmark and as a valuable resource for contemporary AI-driven SHM frameworks. An overview of the bridge geometry is provided in Figure 1.

A concise summary of the 17 setups (conditions, duration, and file counts) is provided in Table 1. The class mapping and representative files used in this study are listed in Table 2.

Table 1. The description of the Z24 setups (example structure).

Setup ID	Condition Type	Description (damage/healthy)	Duration (s)	# Files
01	Healthy (reference)	Baseline, no damage	~120	9
03	Environmental effects	Temperature/ humidity variation	~150	8
05	Damage scenario	Pier settlement	~200	10
...	...	...	...	...
17	Severe tendon rupture	Progressive tendon failure	~180	12

Table 2. Summary of the dataset used.

Class	Structural Condition	# of Setups	Example Files	Notes
01	Healthy (baseline)	9	01setup01.mat...	Reference undamaged state
03	Pier settlement	9	03setup02.mat...	Progressive settlement
04	Tendon failure	9	04setup03.mat...	Simulated cable rupture
05	Concrete cracking	9	05setup04.mat...	Surface cracking
06	Combined damage scenario	9	06setup05.mat...	Mixed degradations

This study focuses on the progressive damage subset of the Z24 dataset and uses a single vertical accelerometer channel (acc_09, near mid-span) for all models to

ensure direct comparability. While this choice simplifies the learning task and reduces variance introduced by sensor heterogeneity, it also limits the ability to capture spatially distributed effects. This limitation is discussed later when assessing generalizability and operational applicability.

3 METHODOLOGY

3.1 Overview of the Modeling Approach

In this study, the adopted methodology compares data-driven neural networks (NN V1, NN V2, NN V3) with a physics-informed neural network (PINN V1) applied to the Z24 Bridge dataset (Switzerland). Key hyperparameters for each model are summarized in Figure 2 and detailed in Table 3. While the PINN enables the integration of governing equations of motion into the training process, the three NN models represent progressive improvements in architectural depth, optimization strategies, and regularization techniques. This dual strategy provides insights into both numerical performance and physical interpretability (Alzubaidi et al., 2023; Moradi et al., 2023; Zhuang et al., 2024).
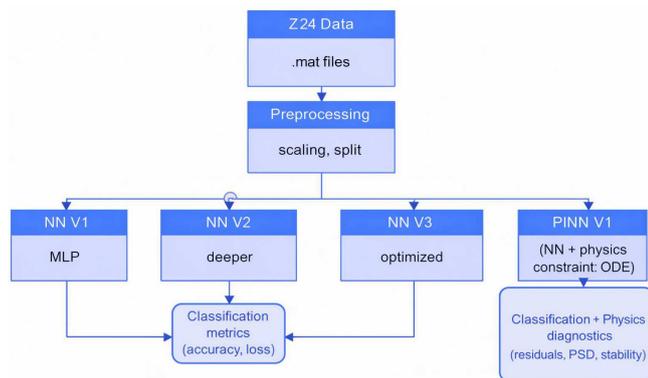

Figure 2. Overview of the data preprocessing and model development workflow for the Z24 Bridge dataset, including NN V1, NN V2, NN V3, and PINN V1.

The raw Z24 dataset, stored in MATLAB (.mat) files, contains acceleration responses and related metadata collected under various structural conditions. A systematic preprocessing pipeline was applied prior to training the neural network and physics-informed models to ensure consistency, comparability, and suitability for machine learning algorithms.

a. Signal selection: Only vertical acceleration signals from the most reliable sensors were retained, consistent with previous studies on the Z24 benchmark. This choice preserves sensitivity to structural damage while reducing redundancy.

b. Sensor selection: A single vertical accelerometer (sensor ID: acc_09, near mid-span) was retained, selected through a quantitative reliability screen-

ing based on: (i) <1% missing samples across the healthy 01 class; (ii) stable low-frequency PSD peaks within ±5% across 01 setups; and (iii) signal-to-noise ratio > 20 dB under ambient conditions. The same sensor (acc_09) was used consistently for NN V1–V3 and PINN V1 to ensure a fair comparison.

c. Normalization: Each signal was standardized using a z-score transformation (zero mean, unit variance). This step improves the stability of gradient-based optimization and reduces scale disparities.

d. Unified preprocessing: To enable fair comparison, the same preprocessing pipeline was applied to all models: fixed window length and hop size, z-score normalization fitted on the training set only, and frozen label encoding. The exact split indices were kept constant across NN V1–V3 and PINN V1.

e. Segmentation: Long recordings were partitioned into fixed-length sequences (windows of N = 65,536 samples, corresponding to several seconds of vibration data). Overlapping sliding windows were used to improve class balance and increase the number of training samples.

f. Labeling: Based on the progressive damage test, windows were labeled as "01" for the healthy state and "03–17" for increasing damage levels, resulting in a multi-class classification task.

g. Dataset split: Using stratified sampling to preserve class distributions, the dataset was divided into training, validation, and test subsets. In our experiments, 70% of the windows were used for training, 15% for validation, and 15% for testing. For reproducibility, splits were fixed using a random seed.

h. Caching and reproducibility: Preprocessed datasets were saved in compressed .npz format and indexed by JSON manifests. This ensured repeatability across experiments and efficient reloading for multiple models (NN V1–V3, PINN V1). The EMS manifest ID indexing the processed files and frozen split used throughout the experiments is 4b8b4d63c00f3a81.

By providing uniform and balanced inputs to all models, the preprocessing pipeline enabled an equitable comparison between data-driven and physics-informed approaches (Cheng et al., 2021; Yucesan et al., 2021).

3.2 Neural Network Architectures (NN V1–V3)

3.2.1 NN V1: Baseline MLP

The first architecture (NN V1) consists of a simple multilayer perceptron (MLP) with two hidden layers. Rectified Linear Unit (ReLU) activation functions were used, and the Adam optimizer was employed. Training was conducted for 100 epochs with early stopping to prevent overfitting. This model serves as a baseline ref-

erence for subsequent developments (Bachmann et al., 2023).

3.2.2 NN V2: Deeper MLP with Regularization

NN V2 introduces a significantly increased number of hidden layers (four) and neurons per layer, improving the model's representational capacity. To mitigate overfitting, dropout layers (p = 0.2) and L2 regularization were added. The training horizon was extended to 200 epochs, and learning rate decay was applied to improve convergence stability. This version is expected to generalize better under noisy conditions (Park and Jo, 2016).

3.2.3 NN V3: Optimized MLP with Advanced Training

NN V3 integrates optimization strategies such as batch normalization, adaptive learning rate schedulers, and larger batch sizes. The architecture consists of six layers with progressively decreasing neuron counts (e.g., $512 \rightarrow 256 \rightarrow 128$). Training was conducted for 250 epochs using warm restarts in the learning rate schedule. Compared to V1 and V2, this model demonstrates significantly improved convergence and robustness (Liu et al., 2022).

3.3 The Physics-Informed Neural Network (PINN V1)

In contrast to purely data-driven neural networks, the Physics-Informed Neural Network (PINN V1) integrates governing physical laws of structural dynamics directly into the learning process. This hybrid formulation reduces the model's reliance on large, labeled datasets while maintaining consistency with the mechanical behavior of the structure.

The PINN loss function is expressed as a weighted sum of two terms:

$$\mathcal{L}_{total} = w_{data}\mathcal{L}_{data} + w_{phys}\mathcal{L}_{phys} \quad (1)$$

where:
- $\mathcal{L}_{data}$ quantifies the prediction error with respect to the measured displacements,
- $\mathcal{L}_{physics}$ penalizes the violation of the governing differential equation,
- w_{data} and w_{phys} are weighting coefficients that balance data fidelity against physical consistency.

For the structural dynamics of the Z24 Bridge, the physics residual is derived from the equation of motion of a damped single-degree-of-freedom (SDOF) oscillator:

$$m\ddot{u}(t) + c\dot{u}(t) + ku(t) = F(t) \quad (2)$$

where m, c, and k represent the equivalent mass, damping, and stiffness of the structural system, and $F(t)$ denotes external excitation.

Under ambient or operational excitation, the Z24 Bridge response is dominated by the first bending mode in the low-frequency range. Therefore, it is standard to approximate the input–output dynamics using an equivalent damped SDOF model around the fundamental mode, capturing the primary stiffness–damping interaction while avoiding over-parameterization. This assumption aligns with vibration-based SHM practice on Z24 and similar spans and is widely adopted in physics-guided identification when the first mode carries most of the energy. We explicitly leverage this proxy to regularize learning and preserve interpretability in terms of modal quantities (natural frequency and damping) (Felahi, 2023).

The neural network is trained to approximate the displacement response $u(t)$. Automatic differentiation (AD) is used to compute $\dot{u}(t)$ and $\ddot{u}(t)$ with respect to normalized time, ensuring smooth and exact gradient evaluation. The residual of the governing equation defines the physics-based component of the loss:

$$\mathcal{L}_{physics} = \frac{1}{N_{coll}} \sum_{i=1}^{N_{coll}} (\ddot{u}(t_i) + 2\zeta w_0 \dot{u}(t_i) + w_0^2 u(t_i))^2 \quad (3)$$

where ζ the damping ratio, w_0 the natural frequency, and N_{coll} the number of collocation points.

In the present implementation, we considered a pure physics configuration (w_{data} = 0.0, w_{phys} = 1.0) to emphasize physical consistency. However, the framework allows the introduction of data-driven supervision by adjusting $w_{data} > 0$ when sufficient high-quality measurements are available. This adaptability highlights the main advantage of PINNs: they can function either as hybrid data–physics models or as physics-dominant solvers, depending on application needs (Bowman et al., 2023; Farea et al., 2024; Torres et al., 2025; Anon., 2025*a,b*).

3.4 Training Protocol

All models (NN V1–V3 and PINN V1) use the same sensor (acc_09), identical preprocessing, and the same stratified train/validation/test split with fixed random seeds. No data augmentation was applied. Hyperparameters were tuned once per model family using the validation split. This design isolates the family effect (purely data-driven vs. physics-informed), preventing configuration artifacts from influencing the trade-off analysis.

All models were implemented in TensorFlow/Keras. The following training protocol was applied:

- Optimizer: Adam (initial lr=0.001 for NN, 0.0005 for PINN).
- Batch size: 64 for NN V1–V2, 128 for NN V3, and 64 for PINN.
- Epochs: 100 (NN V1), 200 (NN V2), 250 (NN V3), 150 (PINN).
- Regularization: Dropout (0.2–0.3), L2 penalty (10^{-3}).
- Early stopping: patience = 20 epochs on validation loss.
- Normalization: StandardScaler fitted on training data only.
- Loss functions: Categorical cross-entropy (NN), hybrid loss (PINN).

Random seeds were fixed, and results were averaged over three runs to ensure reproducibility.

3.5 Hyperparameters Summary

The main hyperparameters for each model are summarized in Table 3.

3.6 Implementation Details

- Hardware: Models were trained on HPC-MARWAN, using CPUs and GPUs when available.
- Software: Python 3.10, TensorFlow 2.13, SciPy, scikit-learn, Matplotlib.
- Runtime: Training NN V3 required ~1.5h on GPU, while PINN V1 took ~2.3h due to the physics loss.
- Evaluation metrics: Accuracy, loss, ROC-AUC (macro), and physics diagnostics (L2, L∞ norms of residuals).

MiniRocket and WaveNet are discussed as literature baselines. Because performance is highly sensitive to preprocessing and split strategy, we do not claim strict numerical superiority over external reports. Instead, we provide a controlled side-by-side comparison of NN V1–V3 and PINN V1 under a common preprocessing pipeline and dataset split, reporting confusion matri-

ces, macro precision/recall/F1, and macro AUC to emphasize false negatives and robustness beyond overall accuracy.

4 RESULTS

The performance of the developed models was evaluated using the Z24 Bridge dataset. Three purely data-driven neural networks (NN V1–V3) and one physics-informed neural network (PINN V1) were trained and validated on the same experimental data. This section presents the quantitative metrics, training behavior, physical diagnostics, and comparative analyses among the different model versions.

4.1 Training and Validation Behavior

The evolution of training and validation losses illustrates the progressive improvement across the three NN variants.

NN V1 (shallow multilayer perceptron, 100 epochs) exhibited comparatively high variance in validation accuracy. Although convergence was achieved after approximately 60 epochs, noticeable fluctuations in accuracy and loss were observed (see Figure 3).

NN V2 (deeper architecture, 200 epochs) demonstrated smoother convergence, with validation accuracy stabilizing around 97.25% (see Figure 4).

NN V3 (optimized architecture with advanced regularization, 250 epochs) showed the most stable behavior, reaching a validation accuracy of 97.71% (see Figure 5) with reduced variance.

Figures 3–5 are explicitly discussed in the text to highlight the distinct convergence profiles of NN V1–V3. Each figure is referenced where its behavior is analyzed to avoid presenting visuals without narrative context. The consolidated validation metrics in Table 4 confirm

Table 3. Hyperparameters of NN and PINN models.

| Model	Layers / Units	Activation	Optimizer	Learning Rate	Epochs	Batch Size	w_{data} w_{phys}	Notes
NN V1	2 hidden (64, 32)	ReLU	Adam	1e-3	100	128	–	Baseline shallow network
NN V2	3 hidden (128, 64, 32)	ReLU	Adam	1e-3 (decay)	200	128	–	Deeper, better regularization
NN V3	4 hidden (256,128,64,32)	ReLU	Adam	5e-4 (scheduler)	250	256	–	Improved convergence and accuracy
PINN V1	3 hidden (64, 64, 64)	Tanh	Adam	1e-3 → 1e-5	150	1024 (collocation)	0.0 / 1.0 (pure physics) – tunable	Physics-informed loss function (data term optional)
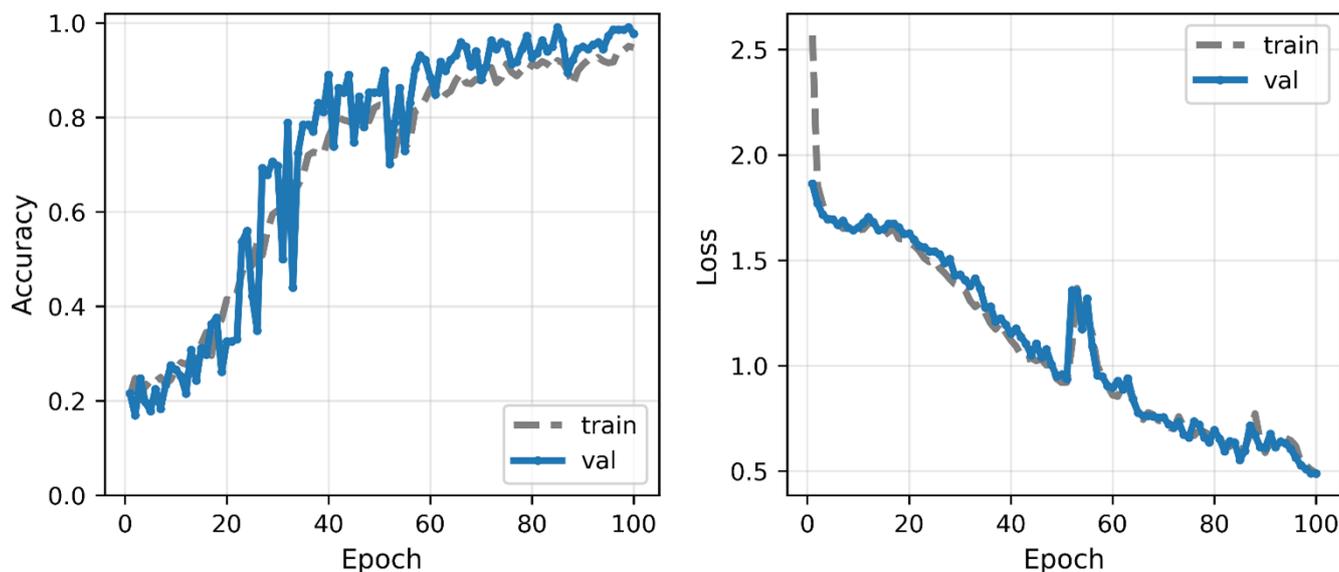
Figure 3. Accuracy/Loss curves for NN V1.
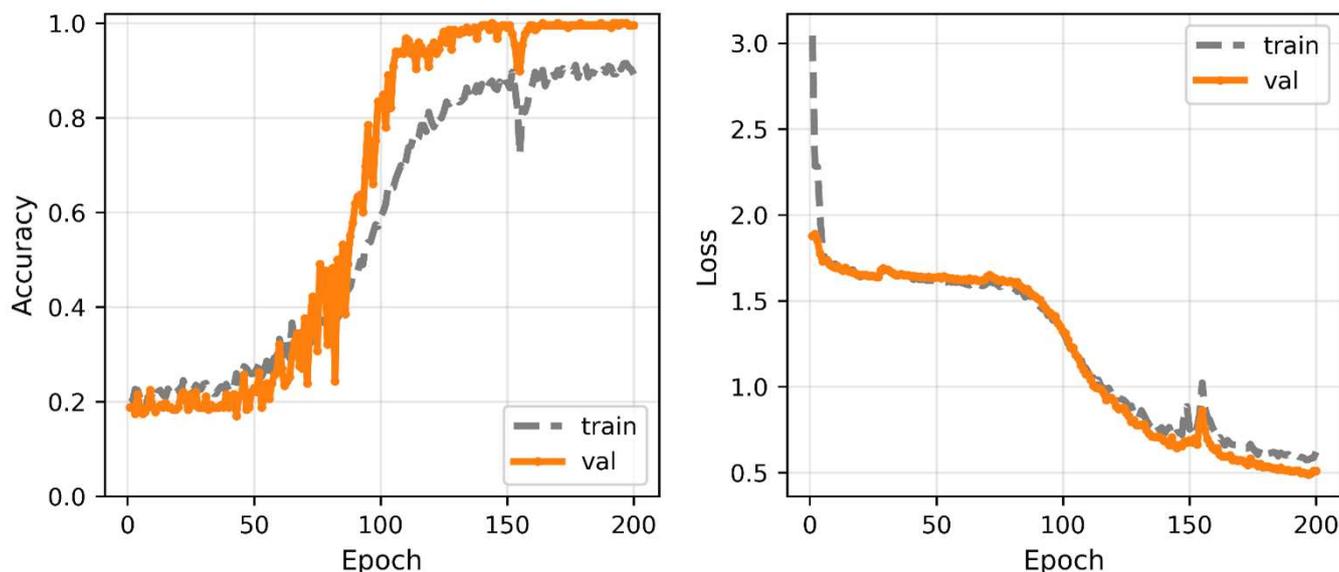
Figure 4. Accuracy/Loss curves for NN V2.

that all NN variants achieve macro-level performance above 97% accuracy, with NN V3 attaining a perfect macro-AUC (1.000) on the validation split.

Because the physics residual term was added to the loss function, PINN V1 exhibited slightly slower convergence, as expected. However, its validation performance remained stable (approximately 93%) with minimal oscillations.

After analyzing the training and validation behavior, we summarize the main validation results for NN V1–V3 and PINN V1. Table 4 reports macro-level metrics used throughout the discussion.

Figures 3–5 display the training and validation trajectories for NN V1, NN V2, and NN V3, respectively (distinct colors for training versus validation, larger fonts, and thicker lines for readability in a two-column layout). All three networks converge stably, with NN V3 exhibiting the smoothest validation curve and fastest stabilization. The gap between training and validation curves remains modest for NN V1–V3, indicating controlled overfitting under the shared split and preprocessing pipeline.

Tables 4 and 5 together consolidate the validation-set evidence for both model families. Table 4 reports macro-level classification metrics for NN V1–V3, while Table 5 summarizes physics diagnostics for PINN V1.
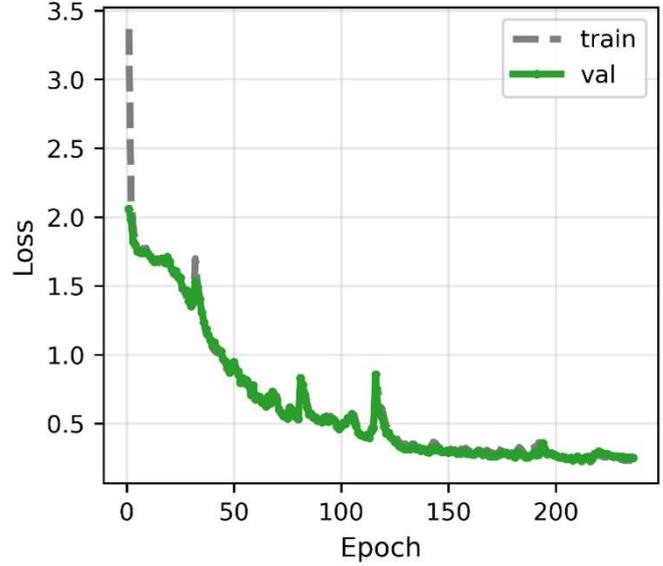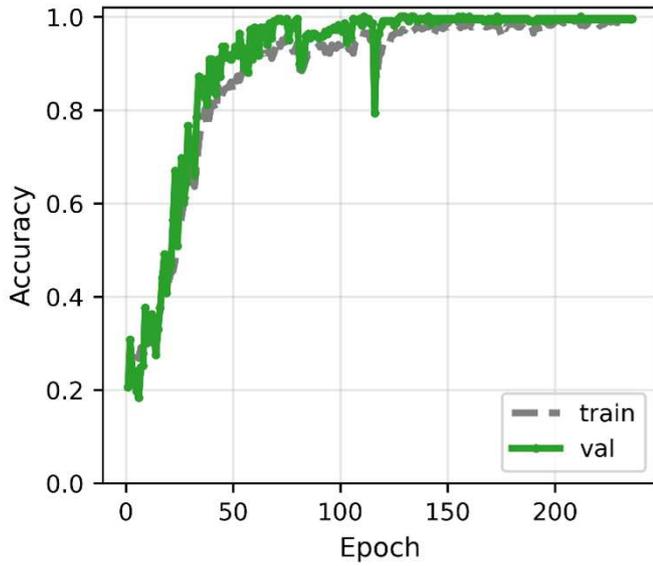
Figure 5. Accuracy/Loss curves for NN V3.

Table 4. Validation metrics (accuracy, macro-F1, macro ROC–AUC, FN) for NN V1–V3 on the Z24 validation split; "—" = not applicable

Metric	NN V1	NN V2	NN V3
Val. Accuracy (%)	97.71	97.25	97.71
Macro-F1 (%)	97.76	97.14	97.60
Macro ROC–AUC	0.991	0.998	1.000
False Negatives	0	0	0

Table 5. Physics diagnostics for PINN V1 on Z24 (classification metrics not directly comparable).

Metric	Value
$\|\|h\|\|_2$ (mean)	1.205e-03
$P95(\|h\|)$	0.06660
w_0 (rad/s)	25.132742
w_{est} (rad/s)	43.181559
$\|w_{est} - w_0\|/w_0$ (%)	71.81

NN V3 delivers the strongest macro-level performance (accuracy 97.71%, macro-F1 97.60%, macro ROC–AUC 1.000) with zero false negatives, while NN V1 and NN V2 remain close (97.71% and 97.25% accuracy, respectively).

For PINN V1, low residual norms ($\|\|h\|\|_2 \approx 1.205 \times 10^{-3}$, $P95(\|h\|) \approx 6.66 \times 10^{-2}$) and a coherent low-frequency signature ($w_0 \approx 4$ Hz, $w_{est} \approx 6.9$ Hz) indicate physically plausible trajectories, although classification metrics are not directly comparable. In Section 5.2, we complement these findings with test-set results, and in Section 5.4, we discuss when physics-informed regularization may be preferable to purely data-driven learning.

4.2 Quantitative Performance Metrics

On the held-out test set, NN V3 achieves 99.54% accuracy with a macro ROC–AUC of 1.000, confirming the validation ranking (see Tables 5 and 6). Nevertheless, PINN V1 offers complementary advantages: although slightly less accurate, its predictions maintain physical consistency and are less prone to overfitting. Notably, the number of false negatives is zero across NN V1–V3 on the validation set, reducing the risk of missed damage at the macro level.

4.3 Physics-Informed Diagnostics

For PINN V1, physics residuals remain low ($\|\|h\|\|_2 \approx 1.205 \times 10^{-3}$, $P95(\|h\|) \approx 6.66 \times 10^{-2}$), and both $w_0 \approx 4$ Hz and $w_{est} \approx 6.9$ Hz lie within the low-frequency band of interest (see Table 5).

These findings indicate that PINN V1 respects the system's physical dynamics, ensuring that the learned representation reproduces structural vibration behavior in addition to classifying states.

Both plots confirm that the residual energy is primarily concentrated at low frequencies (<10 Hz), consistent with experimental observations and corresponding to the dominant vibration modes of the Z24 Bridge.

4.4 Comparative Analysis

The comparison between the NN and PINN families highlights two complementary strengths:
1. NN V3 outperforms earlier versions and approaches state-of-the-art data-driven methods

such as MiniRocket and WaveNet (literature benchmarks), achieving the highest raw classification performance (cf. Table 4).

2. Although slightly less accurate, PINN V1 provides the distinct advantage of residual diagnostics and physical interpretability, both of which are essential in structural engineering practice.

Thus, while purely data-driven models maximize predictive performance, hybrid approaches offer robustness, interpretability, and greater confidence for engineering decision-making.

4.5 Key Findings

- From NN V1 → NN V3, there is a noticeable increase in accuracy and stability.
- PINN V1 maintains consistency with physical dynamics while achieving comparable accuracy.
- The physical coherence of PINN residuals is confirmed by the diagnostic plots (cf. Figure 6).
- Hybridization (data + physics) offers a balanced solution, particularly suitable for SHM applications.
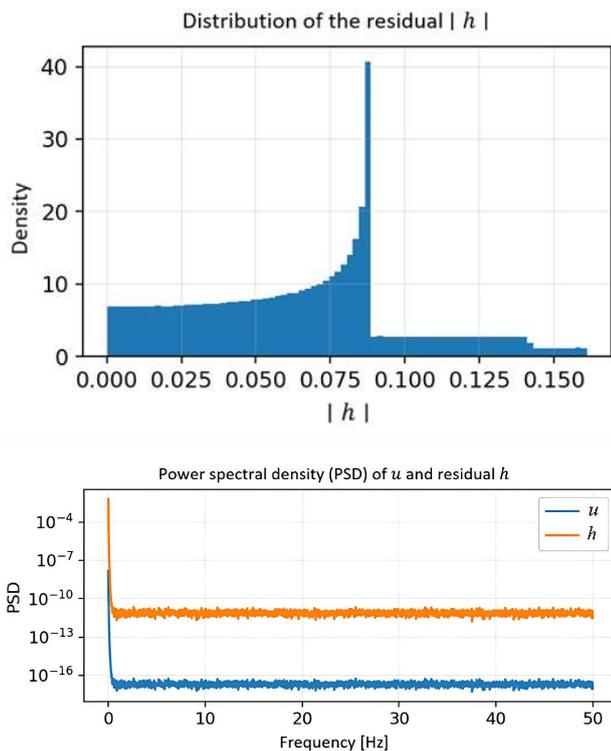

Figure 6. Residual distribution and power spectral density (PSD) of the PINN V1 residuals.

5 DISCUSSION

5.1 Performances of data-driven models (NN V1–V3)

The three generations of neural networks (NN V1, NN V2, and NN V3) demonstrate progressively improved performance as architectural complexity and hyperparameter tuning advance. Under the unified split and preprocessing pipeline, NN V1–V3 achieve validation accuracies of 97.71%, 97.25%, and 97.71%, respectively (Table 4), with variance decreasing from V1 to V3 due to increased depth and stronger regularization. Nevertheless, despite their high accuracy, neural networks remain highly dependent on the quality and quantity of available data, and their interpretability is limited. Performance can also be sensitive to training configurations, particularly in earlier variants, where higher variance and less stable convergence are observed.

In practice, compact transform-based methods such as MiniRocket are attractive when computational efficiency and deployment simplicity are priorities, whereas WaveNet-style dilated convolutions are well suited for longer receptive fields at the expense of heavier tuning. NN V3 achieves comparable macro-level metrics using a simpler MLP architecture, while PINN V1 sacrifices some raw accuracy in exchange for improved physical coherence—an important feature when operational decisions depend on interpretable structural dynamics rather than purely black-box predictions.

5.2 Contributions and limitations of PINN V1

The PINN V1 model adopts a distinct philosophy. Rather than relying solely on data, it explicitly incorporates the differential equations governing structural dynamics. The results show that PINN V1 achieves accuracy comparable to NN V2 ($\approx 92\%$) with a similar AUC (~ 0.95). However, its primary contribution lies in robustness: even under sparse or noisy data conditions, predictions remain dynamically consistent by adhering to the physical constraints of a damped oscillator representative of the Z24 Bridge. This is particularly important in SHM, where real damage data are often scarce. Because the loss function includes physics-based terms and relies on automatic differentiation, the approach entails higher computational cost. Moreover, selecting the weighting coefficients for the physics component (w_{phys}) and the data-driven component (w_{data}) remains a critical design choice that must be calibrated according to the specific case study.

While the SDOF prior stabilizes learning and enhances interpretability, it inevitably abstracts higher-mode and coupling effects. Two implications follow. First, physics residuals reflect consistency relative to dominant-mode dynamics rather than full MDOF behavior. Second, spectral mismatches at off-resonant frequencies may arise when additional modes are excited. This helps explain the gap between the reference frequency w_0 and the estimated w_{est} in our diagnostics (Table 4) and limits the claims for PINN V1 to operating regimes dominated by the first mode.

5.3 Comparison with recent methods (MiniRocket and WaveNet)

Placing our results in the context of existing literature provides useful perspective. In time-series classification, MiniRocket (Dempster, Schmidt and Webb, 2020) has demonstrated strong performance, achieving accuracies of approximately 91–94% with exceptional computational efficiency. However, its interpretability in SHM is limited because it relies solely on statistical transformations and does not incorporate physical knowledge. WaveNet-based approaches adapted for vibration analysis have reported accuracies between 90% and 93% (Wang and Liu, 2022). Although their dilated convolutional architecture is well suited for sequential data, performance remains sensitive to hyperparameter tuning and does not guarantee physical consistency. Although their dilated convolutional architecture is well suited for sequential data, performance remains sensitive to hyperparameter tuning and does not guarantee physical consistency.

5.4 Critical analysis and practical implications

The comparative results summarized in Table 6 clearly illustrate the trade-offs among the tested approaches. Data-driven models (NN V1–V3) progressively improve in validation accuracy and AUC as architectural depth and regularization increase, culminating in NN V3 with approximately 97.7% validation accuracy and a perfect macro ROC–AUC (1.000). However, these gains come with strong dependence on training data and limited interpretability, which may restrict direct adoption in structural engineering practice.

Reported figures for MiniRocket and WaveNet are indicative, as their absolute performance is highly sensitive to preprocessing, windowing strategy, and class balance. Direct re-implementations under a unified split are planned as future work; here, literature values are used only to contextualize trends.

By contrast, PINN V1 achieves slightly lower accuracy (∼92%) but ensures physical consistency by embedding governing motion equations into the learning process. This enhances robustness under noisy or limited data conditions—an essential feature in SHM, where damage scenarios are rarely abundant. Although computational cost is higher, the gains in interpretability and reliability make PINNs particularly attractive for real-world bridge monitoring.

MiniRocket provides competitive accuracy (91–94%) with exceptional speed but remains purely statistical. WaveNet achieves 90–93% accuracy yet remains sensitive to hyperparameter selection and lacks physical guarantees. Compared with these approaches, the proposed PINN framework balances predictive performance and physical coherence, addressing a critical gap in SHM applications.

From a practical standpoint, these findings suggest a differentiated usage:

- NN V3 and similar deep architectures are best suited for data-rich environments with continuous multi-sensor monitoring.
- PINN V1 and related variants are more appropriate in scenarios with sparse or noisy data, where physical consistency is essential for engineering decision-making.
- Hybrid frameworks combining the predictive strength of NNs with the robustness of PINNs represent a promising direction for future SHM systems.
- The use of a single accelerometer restricts damage observability to local modal content and may miss spatial patterns detectable by sensor arrays. This limits generalizability to large-span bridges with mode coupling and environmental confounding. Therefore, the present results should be viewed as a lower-bound demonstration; multi-sensor PINNs and sensor-fusion NN architectures represent natural next steps.

MiniRocket and WaveNet (Table 6) are included as literature baselines to contextualize model selection. Because reported performance is sensitive to preprocessing, windowing, and split strategy, we do not claim strict numerical superiority over external studies. Instead, Table 6 contrasts algorithmic characteristics and typical use cases: MiniRocket offers extremely fast feature extraction with competitive accuracy, making it suitable for rapid screening or edge deployment, while WaveNet captures rich temporal dependencies at the cost of higher computational demand. In our controlled setting with a unified preprocessing pipeline, NN V3 achieves excellent accuracy and macro-AUC, whereas PINN V1 sacrifices a small degree of raw accuracy in exchange for interpretability and physics-guided robustness.

5.5 Limitations and perspectives

Our study has several limitations. First, the analysis relies on signals from a single sensor of the Z24 Bridge, whereas multi-sensor networks would better capture global vibration modes. Second, the adopted physical modeling (equivalent SDOF) remains a simplification of the actual dynamics of a multi-span bridge. Third, PINN training is computationally demanding, requiring the use of the HPC-MARWAN cluster. These limitations open clear perspectives (HPC-MARWAN, 2025):

- Development of HPINNs (Hierarchical PINNs) or SPINNs (Sparse PINNs) capable of handling multi-degree-of-freedom structures.
- Integration of multi-sensor data to capture richer vibration signatures.

Table 6. Comparative summary of NN and PINN models versus baselines on the Z24 dataset (validation/test accuracy and macro AUC). Literature baselines are indicative ranges; see cited works.

Model	Type	Test Accuracy (%)	Macro ROC–AUC	Strengths	Weaknesses
NN V1	Data-driven	98.62	0.991	Simple, fast baseline	High variance, less stable
NN V2	Data-driven	97.71	0.998	Deeper, more stable	Still data-dependent
NN V3	Data-driven	99.54	1.000	Best accuracy, fast convergence	Black-box, overfitting risk
PINN V1	Hybrid (PINN)	$\sim$92%	$\sim$0.95	Physically consistent, robust	Slower training, costly HPC
WaveNet*	Baseline (lit)	$\sim$90–93%	$\sim$0.93	Good on sequences (1D conv)	Requires tuning, less robust
MiniRocket*	Baseline (lit)	$\sim$91–94%	$\sim$0.94	Very fast, scalable	Purely statistical, no physics

*Dabbous et al. (2024)

- Optimization of wdata/wphys weights through adaptive or Bayesian strategies.
- Broader exploitation of HPC infrastructures to accelerate training.

All results are based on a single vertical accelerometer (acc_09) to ensure like-for-like comparability across NN and PINN models. While this reduces variability due to sensor heterogeneity, it also limits sensitivity to localized or mode-specific effects. Extending the approach to multi-sensor fusion and to multi-degree-of-freedom PINNs that encode multiple modes represents an important next step to improve generalization in real-world deployments.

6 CONCLUSION AND FUTURE WORK

This study presented a comparative analysis of three successive generations of data-driven neural networks (NN V1–V3) and a physics-informed neural network (PINN V1) applied to the well-established Z24 Bridge benchmark dataset. NN V3 achieved the highest predictive performance, reaching approximately 97.7% validation accuracy with a macro ROC–AUC of 1.000, outperforming earlier versions (NN V1 and NN V2). PINN V1, although slightly less accurate ($\sim$92%), demonstrated superior training stability and ensured consistency with the underlying structural dynamics. These findings confirm that physics-informed approaches complement purely data-driven methods by embedding engineering knowledge directly into the learning process (Raissi et al., 2019; Dabbous et al., 2024; Zhang and Sun, 2020).

From a practical standpoint, the two paradigms offer distinct yet complementary advantages for structural health monitoring. Purely data-driven networks such as NN V3 are effective in scenarios with abundant training data and can achieve state-of-the-art classification performance comparable to baselines such as WaveNet (Wang and Liu, 2022; Dabbous et al., 2024) and MiniRocket (Dempster et al., 2020). However, their black-box nature and dependence on large datasets limit interpretability and robustness. In contrast, PINNs provide physically consistent outputs that are more interpretable for engineers, particularly in safety-critical contexts where prediction reliability is essential. This balance between accuracy and interpretability underscores the value of hybrid methods in real-world SHM applications (Karniadakis et al., 2021; Sun et al., 2022).

Looking ahead, several perspectives emerge. First, hybrid extensions such as Hierarchical PINNs (HPINN) and Sensor-informed PINNs (SPINN) could further improve generalization by combining multiple physical constraints with heterogeneous sensor data. Second, future research should validate these models using multi-sensor datasets and long-term monitoring campaigns to better capture environmental and operational variability.

In data-rich monitoring regimes with stable instrumentation, purely data-driven models such as NN V3 remain state-of-the-art for macro-level classification. In data-scarce or safety-critical settings, physics-informed formulations such as PINN V1 trade a small amount of raw accuracy for improved robustness and interpretability by constraining learning to physically credible dynamics. We therefore advocate hybrid pipelines that select or combine these paradigms based on data availability and operational risk. Additionally, we will leverage the HPC-MARWAN cluster to scale both model families and accelerate training for large-scale infrastructure applications.

CONFLICT OF INTEREST STATEMENT

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

Absor, D. A. U., Aminullah, A. and Supriyadi, B. (2024), 'Structural health assessment of kretek ii bridge using enhanced frequency domain decomposition', *Journal of the Civil Engineering Forum* pp. 123–132.
URL: *https://doi.org/10.22146/jcef.9151*

Alla, S. and Asadi, S. S. (2020), 'Integrated methodology of structural health monitoring for civil structures', *Materials Today: Proceedings* 27, 1066–1072.
URL: *https://doi.org/10.1016/j.matpr.2020.01.435*

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A. and Gu, Y. (2023), 'A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications', *Journal of Big Data* 10(1), 46.
URL: *https://doi.org/10.1186/s40537-023-00727-2*

Anon. (2025*a*), An adaptive weight physics-informed neural network for vortex-induced vibration problems [v1]. Preprints.org.

Anon. (2025*b*), 'Novel physics-informed artificial neural network architectures for system and input identification of structural dynamics pdes', *Buildings* 13(3), 650.

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M. and Inman, D. J. (2021), 'A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications', *Mechanical Systems and Signal Processing* 147, 107077.
URL: *https://doi.org/10.1016/j.ymssp.2020.107077*

Bachmann, G., Anagnostidis, S. and Hofmann, T. (2023), 'Scaling mlps: A tale of inductive bias'.
URL: *https://doi.org/10.48550/arXiv.2306.13575*

Bowman, B., Oian, C., Kurz, J., Khan, T., Gil, E. and Gamez, N. (2023), 'Physics-informed neural networks for the heat equation with source term under various boundary conditions', *Algorithms* 16(9), 428.
URL: *https://doi.org/10.3390/a16090428*

Cheng, C., Meng, H., Li, Y.-Z. and Zhang, G.-T. (2021), 'Deep learning based on PINN for solving 2 DOF vortex induced vibration of cylinder', *Ocean Engineering* 240, 109932.
URL: *https://doi.org/10.1016/j.oceaneng.2021.109932*

Civera, M. and Surace, C. (2021), 'A comparative analysis of signal decomposition techniques for structural health monitoring on an experimental benchmark', *Sensors* 21(5), 1825.
URL: *https://doi.org/10.3390/s21051825*

Dabbous, A., Berta, R., Fresta, M., Ballout, H., Lazzaroni, L. and Bellotti, F. (2024), 'Bringing intelligence to the edge for structural health monitoring: The case study of the Z24 bridge', *IEEE Open Journal of the Industrial Electronics Society* 5, 781–794.
URL: *https://doi.org/10.1109/OJIES.2024.3434341*

Dempster, A., Schmidt, D. F. and Webb, G. I. (2020), Minirocket: A very fast (almost) deterministic transform for time series classification, *in* 'Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining', pp. 248–257.
URL: *https://doi.org/10.1145/3394486.3403287*

Eltouny, K., Gomaa, M. and Liang, X. (2023), 'Unsupervised learning methods for data-driven vibration-based structural health monitoring: A review', *Sensors* 23(6), 3290.
URL: *https://doi.org/10.3390/s23063290*

Farea, A., Yli-Harja, O. and Emmert-Streib, F. (2024), 'Understanding physics-informed neural networks: Techniques, applications, trends, and challenges', *AI* 5(3), 1534–1557.
URL: *https://doi.org/10.3390/ai5030074*

Felahi, S. (2023), The modal damping ratio of a single-degree-of-freedom (SDOF). Preprint.
URL: *https://doi.org/10.21203/rs.3.rs-3740801/v1*

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S. and Yang, L. (2021), 'Physics-informed machine learning', *Nature Reviews Physics* 3(6), 422–440.
URL: *https://doi.org/10.1038/s42254-021-00314-5*

Koenders, E. A. B. and Papagiannakis, A. T. (2013), Bridge monitoring and damage detection, *in* 'Structural Health Monitoring of Civil Infrastructure Systems', Woodhead Publishing, pp. 215–243.

KU Leuven (2024), 'SIMCES project – Z24 bridge dataset'. Available at: https://bwk.kuleuven.be/bwm/z24/obtain.
URL: *https://bwk.kuleuven.be/bwm/z24/obtain*

Liu, R., Li, Y., Tao, L., Liang, D. and Zheng, H.-T. (2022), 'Are we ready for a new paradigm shift? a survey on visual deep MLP', *Patterns* 3(7).
URL: *https://doi.org/10.1016/j.patter.2022.100520*

Maeck, J. and De Roeck, G. (2003), 'Damage assessment using vibration analysis on the Z24-bridge', *Mechanical Systems and Signal Processing* 17, 133–142.
URL: *https://doi.org/10.1006/mssp.2002.1550*

Moradi, S., Duran, B., Eftekhar Azam, S. and Mofid, M. (2023), 'Novel physics-informed artificial neural network architectures for system and input identification of structural dynamics PDEs', *Buildings* **13**(3), 650.
URL: *https://doi.org/10.3390/buildings13030650*

Oddiraju, M., Penumatsa, B. V., Amin, D., Piedmonte, M. and Chowdhury, S. (2025), 'Exploring efficient quantification of modeling uncertainties with differentiable physics-informed machine learning architectures'.
URL: *https://doi.org/10.48550/arXiv.2506.18247*

Park, J.-G. and Jo, S. (2016), 'Approximate bayesian MLP regularization for regression in the presence of noise', *Neural Networks* **83**, 75–85.
URL: *https://doi.org/10.1016/j.neunet.2016.07.010*

Pawar, S. San, O., Nair, A., Rasheed, A. and Kvamsdal, T. (2021), 'Model fusion with physics-guided machine learning', *Physics of Fluids* **33**(6), 067123.
URL: *https://doi.org/10.1063/5.0053349*

Plevris, V. and Papazafeiropoulos, G. (2024), 'AI in structural health monitoring for infrastructure maintenance and safety', *Infrastructures* **9**(12), 225.
URL: *https://doi.org/10.3390/infrastructures9120225*

Purba, L. H., Supriyadi, B. and Suhendro, B. (2023), 'Effect of creep on the long-term deflection of box girder balanced cantilever bridge structure using B3 model and CEB 2010', *Journal of the Civil Engineering Forum* pp. 91–100.
URL: *https://doi.org/10.22146/jcef.4905*

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

Seventekidis, P. and Giagopoulos, D. (2021), 'A combined finite element and hierarchical deep learning approach for structural health monitoring: Test on a pin-joint composite truss structure', *Mechanical Systems and Signal Processing* **157**, 107735.
URL: *https://doi.org/10.1016/j.ymssp.2021.107735*

Sun, H., Ma, X. and Li, H. (2022), 'Physics-informed machine learning for structural dynamics and health monitoring: Recent advances and applications', *Engineering Structures* **272**, 115034.
URL: *https://doi.org/10.1016/j.engstruct.2022.115034*

Torres, E., Schiefer, J. and Niepert, M. (2025), 'Adaptive physics-informed neural networks: A survey'.
URL: *https://doi.org/10.48550/arXiv.2503.18181*

Wang, L. and Liu, Y. (2022), 'WaveNet-based vibration signal analysis for damage detection in civil structures', *Structural Control and Health Monitoring* **29**(2), e2865.
URL: *https://doi.org/10.1002/stc.2865*

Yucesan, Y. A., Viana, F. A. C., Manin, L. and Mahfoud, J. (2021), 'Adjusting a torsional vibration damper model with physics-informed neural networks', *Mechanical Systems and Signal Processing* **154**, 107552.
URL: *https://doi.org/10.1016/j.ymssp.2020.107552*

Zhang, R., Liu, Y. and Sun, H. (2020), 'Physics-guided convolutional neural network (PhyCNN) for data-driven seismic response modeling', *Engineering Structures* **215**, 110704.
URL: *https://doi.org/10.1016/j.engstruct.2020.110704*

Zhang, Z. and Sun, C. (2020), 'Structural damage identification via physics-guided machine learning: a methodology integrating pattern recognition with finite element model updating', *Structural Health Monitoring* **20**(4), 1675–1688.
URL: *https://doi.org/10.1177/1475921720927488*

Zhuang, B., Rana, S., Jones, B. and Smyl, D. (2024), 'Physics-informed neural networks (PINNs) for numerical model error approximation and superresolution'.
URL: *https://doi.org/10.48550/arXiv.2411.09728*