Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Spectral bias mitigation in physics-informed neural surrogates for nonlinear structural dynamics: Fourier feature encoding versus Kolmogorov-Arnold representations

Salih Berkan Aydemir

salih.aydemir@amasya.edu.tr

Amasya University

Research Article

Keywords:

Posted Date: July 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9643432/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Spectral bias mitigation in physics-informed neural surrogates for nonlinear structural dynamics: Fourier feature encoding versus Kolmogorov–Arnold representations

Salih Berkan Aydemir[1*]

[1*]Computer Engineering, Amasya University, Amasya, TURKEY.

Corresponding author(s). E-mail(s): salih.aydemir@amasya.edu.tr;

Abstract

Standard multilayer perceptrons (MLPs) employed in physics-informed neural networks (PINNs) suffer from spectral bias—the systematic inability to learn high-frequency components in multi-frequency structural response signals. Recently, Kolmogorov–Arnold Networks (KANs) with learnable edge activations have been proposed as alternatives that could inherently mitigate this limitation. This study presents a rigorous, parameter-matched (∼16,400 parameters per model) comparison of four neural architectures—Fourier-enhanced KAN (FPIKAN), Fourier-MLP, plain MLP, and standalone KAN—across four canonical nonlinear dynamics benchmarks: the Bouc–Wen hysteretic oscillator, the Duffing oscillator with cubic stiffness, the forced Van der Pol oscillator, and a three-degree-of-freedom shear building. A systematic $\mathbf{2 \times 2}$ factorial ablation study isolates the individual contributions of (i) Fourier feature encoding and (ii) backbone architecture choice. Results demonstrate unambiguously that Fourier feature encoding is the dominant factor, reducing relative $\boldsymbol{L_2}$ errors by 98.6–100% when applied to MLP across all benchmarks. Fourier-MLP achieves the lowest error in all four benchmarks ($\boldsymbol{L_2 \leq 0.014}$, with $\boldsymbol{L_2 < 10^{-3}}$ for Duffing and Van der Pol). Without Fourier encoding, KAN outperforms MLP by up to $\mathbf{3.5\times}$ (Van der Pol: $\boldsymbol{L_2 = 0.164}$ vs. $\mathbf{0.574}$) due to its learnable Chebyshev activations acting as implicit spectral decomposers. Paradoxically, adding Fourier encoding to KAN *degrades* performance in two benchmarks (Duffing: $\mathbf{-201.7\%}$; Van der Pol: $\mathbf{-9.0\%}$), and FPIKAN consistently underperforms Fourier-MLP. Analysis of gradient norms through Chebyshev polynomial layers reveals oscillatory gradient magnitudes incompatible with the smooth optimization landscape required

by second-order physics residual terms. These findings provide critical, experimentally validated guidance for selecting neural architectures in computational structural dynamics.

Keywords: keyword1, Keyword2, Keyword3, Keyword4

1 Introduction

The prediction of dynamic response for structural systems subjected to complex, time-varying loads constitutes one of the central problems in civil and mechanical engineering [1]. Classical time-stepping algorithms—including the Newmark-β family [2], Wilson-θ methods, and explicit Runge Kutta integrators provide highly accurate reference solutions but become computationally prohibitive when repeated evaluations are required, such as in parametric studies, Monte Carlo-based reliability analysis, real-time structural health monitoring, and digital twin frameworks [3].

Physics-informed neural networks (PINNs), introduced by Raissi et al. [4], have emerged as a promising paradigm for constructing data-efficient surrogate models. By embedding governing differential equations directly into the training loss function through automatic differentiation, PINNs produce predictions that satisfy underlying physical laws even when training data is sparse. However, standard PINNs employing multilayer perceptrons (MLPs) with fixed activation functions suffer from *spectral bias*—the well-documented tendency of neural networks to preferentially learn low-frequency components while systematically under-representing high-frequency content [5, 6]. Rahaman et al. [5] established this phenomenon theoretically through neural tangent kernel (NTK) analysis, demonstrating that standard networks converge exponentially faster for low-frequency modes.

This limitation is acutely problematic in structural dynamics, where response signals inherently contain multiple frequency components arising from different vibration modes. A three-story building, for instance, simultaneously vibrates at its fundamental frequency ($\sim$2 Hz) and higher modes ($\sim$6 Hz, $\sim$9 Hz), producing a complex multi-frequency response that standard PINNs fail to capture accurately. Nonlinear systems compound this challenge through subharmonic generation, period-doubling, and energy transfer across frequencies.

Two principal strategies have been proposed. The first is *Fourier feature encoding*, introduced by Tancik et al. [7], which maps scalar inputs to a high-dimensional sinusoidal representation. Wang et al. [8] demonstrated its theoretical basis within PINNs, showing that Fourier features modify the NTK eigenspectrum to equalize convergence rates across frequencies. The second approach involves *Kolmogorov–Arnold Networks* (KANs), introduced by Liu et al. [9], which place learnable univariate functions on network edges rather than fixed activations on nodes, drawing motivation from the Kolmogorov–Arnold representation theorem.

Since the introduction of KANs, several groups have explored physics-informed variants. Shukla et al. [10] presented an early comparison between MLP and KAN for differential equations, while Wang et al. [11] formalized the PIKAN framework. Tao

et al. [12] recently applied Fourier-enhanced PIKAN (FPIKAN) to seismic response prediction, but only for linear elastic structures, leaving nonlinear hysteretic behavior unaddressed.

Despite this growing body of work, a fundamental gap persists: *no study has conducted a rigorously fair comparison between KAN and MLP architectures under controlled conditions*—with matched parameter counts, proper input normalization, correct initial conditions for second-order systems, and systematic ablation of Fourier encoding. Prior comparisons are confounded by several issues: unequal parameter counts, unnormalized time inputs causing tanh saturation, and initial condition constraints hardcoded to zero even for nonzero initial states. These flaws produce misleading conclusions.

The present study addresses this gap with five contributions: (1) a parameter-matched (∼16,400 each) comparison across four benchmarks; (2) a 2×2 factorial ablation isolating Fourier encoding and backbone effects; (3) genuine physics-informed training with proper IC and velocity IC enforcement; (4) mathematical analysis of gradient incompatibility between Chebyshev-basis KAN and physics loss; (5) frequency-domain spectral fidelity analysis.

The remainder is organized as follows. Section 2 presents the benchmark problems. Section 3 describes architectures and training. Section 4 presents results. Section 5 concludes.

2 Benchmark Problems

Four canonical problems are selected, each representing a fundamentally different class of dynamic behavior: hysteretic energy dissipation (Bouc–Wen), cubic geometric nonlinearity (Duffing), self-excited limit-cycle oscillation (Van der Pol), and linear multi-frequency modal superposition (3-DOF building). All reference solutions are computed with stringent numerical tolerances.

2.1 Benchmark 1: Bouc–Wen hysteretic oscillator

The Bouc–Wen model [13], refined by Baber and Noori [14], is the most widely adopted phenomenological model for rate-independent hysteretic behavior in structural engineering, with over 4,000 citations. It captures yielding, plastic deformation, hysteretic energy dissipation, and stiffness degradation in materials such as reinforced concrete and structural steel [15, 16]. The governing equations for an SDOF system under base excitation $\ddot{u}_g(t)$ are:

$$m\ddot{u} + c\dot{u} + \alpha k u + (1-\alpha)kz = -m\ddot{u}_g(t) \tag{1}$$

where the hysteretic variable z evolves according to:

$$\dot{z} = A\dot{u} - \beta|\dot{u}||z|^{n-1}z - \gamma_{bw}\dot{u}|z|^n \tag{2}$$

Here u is relative displacement, $m = 1.0$ kg, $\omega_n = 2\pi/T$ with $T = 0.5$ s (typical low-rise RC frame), $k = m\omega_n^2 = 157.91$ N/m, $c = 2\zeta m\omega_n$ with $\zeta = 0.05$, $\alpha = 0.05$

(post-yield stiffness ratio), and $A = 1$, $\beta = \gamma_{bw} = 0.5$, $n = 1$ following Foliente [15]. The excitation is a deterministic multi-harmonic ground acceleration with six components spanning 0.8–10 Hz, modulated by a Jennings envelope. Reference: RK45, $rtol = 10^{-10}$, $\Delta t = 0.01$ s, $T_{total} = 25$ s, 2,500 points. IC: $u(0) = \dot{u}(0) = z(0) = 0$.

Remark 1 The coupled hidden variable z in Eq. (2) is not directly observable and would require multi-output architecture to embed as physics loss. This benchmark uses data-driven training only, representing the realistic scenario of learning from sensor measurements without internal state access.

2.2 Benchmark 2: Duffing oscillator

The Duffing equation [17–19] models cubic geometric stiffness, with applications in buckled beams, MEMS resonators [20], and energy harvesting:

$$\ddot{x} + \delta\dot{x} + \alpha_d x + \beta_d x^3 = \gamma_d \cos(\omega_d t) \tag{3}$$

with $\delta = 0.3$, $\alpha_d = -1$ (double-well potential), $\beta_d = 1$, $\gamma_d = 0.5$, $\omega_d = 1.2$ rad/s (quasi-periodic regime). IC: $x(0) = 0.5$, $\dot{x}(0) = 0$. Reference: RK45, $rtol = 10^{-10}$, $\Delta t = 0.02$ s, $T_{total} = 60$ s, 3,000 points. Physics-informed training enabled.

2.3 Benchmark 3: Forced Van der Pol oscillator

The Van der Pol oscillator [21] models self-excited vibrations from negative damping, with applications in aeroelastic flutter, cardiac rhythms, and stick-slip friction [22, 23]:

$$\ddot{x} - \mu(1 - x^2)\dot{x} + x = A_f \cos(\omega_f t) \tag{4}$$

with $\mu = 3$ (strongly nonlinear relaxation oscillation), $A_f = 5$, $\omega_f = 1.0$ rad/s. The term $-\mu(1 - x^2)\dot{x}$ injects energy for $|x| < 1$ and dissipates for $|x| > 1$. IC: $x(0) = 0.1$, $\dot{x}(0) = 0$. Reference: RK45, $rtol = 10^{-10}$, $\Delta t = 0.02$ s, $T_{total} = 50$ s, 2,500 points. Physics-informed training enabled.

2.4 Benchmark 4: Three-DOF shear building

A three-story shear building [1] represents MDOF structural dynamics:

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = -\mathbf{M}\mathbf{r}\ddot{u}_g(t) \tag{5}$$

with $\mathbf{M} = \text{diag}(1000, 1000, 1000)$ kg, $\mathbf{r} = [1, 1, 1]^T$, inter-story stiffnesses $k_1 = 2.0\times10^6$, $k_2 = 1.8 \times 10^6$, $k_3 = 1.5 \times 10^6$ N/m, Rayleigh damping with $\zeta = 0.05$ in the first two modes, and eight-harmonic deterministic excitation spanning 1–12 Hz. Top-floor displacement $u_3(t)$ is the target. Reference: Newmark-β ($\tilde{\gamma} = 0.5$, $\tilde{\beta} = 0.25$), $\Delta t = 0.005$ s, $T_{total} = 15$ s, 3,000 points. IC: $\mathbf{u}(0) = \dot{\mathbf{u}}(0) = \mathbf{0}$. Data-driven training only (matrix ODE).

Table 1 summarizes all benchmarks.

Table 1 Summary of benchmark characteristics.

Property	Bouc–Wen	Duffing	Van der Pol	3-DOF
Nonlinearity	Hysteretic	Cubic	Self-excited	Linear
T_{total} (s)	25	60	50	15
Data points	2,500	3,000	2,500	3,000
$x(0)$	0	0.5	0.1	0
Physics loss	No	Yes	Yes	No

3 Methodology

To rigorously evaluate the impact of spectral bias mitigation strategies in predicting nonlinear structural dynamics, this study establishes a unified surrogate modeling framework. The proposed methodology is specifically designed to isolate and independently evaluate two critical architectural decisions: (i) the input representation space (standard normalized time versus high-dimensional Fourier feature encoding) and (ii) the functional backbone of the neural network (classical MLPs with fixed activations versus KANs with learnable Chebyshev polynomials).

To ensure a strictly fair and unbiased comparison, all four evaluated architectures?FPIKAN, Fourier-MLP, plain MLP, and standalone KAN?are constrained to an equivalent model capacity of approximately 16,400 trainable parameters. The subsequent subsections detail the mathematical formulations of the input encodings (Section 3.1), the neural layer operations (Sections 3.2 and 3.3), the unified multi-objective loss function encompassing both data-driven and physics-informed constraints (Section 3.4), and a theoretical analysis of the gradient flows governing their optimization landscapes (Section 3.5).

3.1 Fourier feature encoding

The scalar time input t is mapped to a high-dimensional representation [7, 8]:

$$\Phi(t) = [\cos(2\pi\sigma_1 t),\ \sin(2\pi\sigma_1 t),\ \ldots,\ \cos(2\pi\sigma_K t),\ \sin(2\pi\sigma_K t)] \in \mathbb{R}^{2K} \tag{6}$$

with $K = 20$ log-spaced frequencies: $\sigma_k = \sigma_{\min} \cdot (\sigma_{\max}/\sigma_{\min})^{(k-1)/(K-1)}$. Frequency bounds are adapted per benchmark (Table 2).

Table 2 Fourier frequency ranges.

Benchmark	$\sigma_{\min}$ (Hz)	$\sigma_{\max}$ (Hz)	Rationale
Bouc–Wen	0.3	15.0	Earthquake band
Duffing	0.05	5.0	Low-freq. quasi-periodic
Van der Pol	0.05	5.0	Relaxation oscillation
3-DOF Building	0.5	15.0	Structural modes

The theoretical justification follows from NTK analysis [8]: Fourier features modify the kernel eigenspectrum to equalize convergence rates across frequencies, eliminating the exponential slowdown for high-frequency modes that causes spectral bias.

3.2 KAN layer with Chebyshev basis

The Kolmogorov–Arnold representation theorem states that any continuous $f : [0,1]^n \to \mathbb{R}$ can be written as $f(\mathbf{x}) = \sum_{q=0}^{2n} \Psi_q(\sum_{p=1}^{n} \varphi_{q,p}(x_p))$ with univariate $\varphi_{q,p}$ and Ψ_q [9]. KAN layers implement this via Chebyshev polynomials of degree $D = 5$:

$$\mathrm{KAN}(\mathbf{x}) = \underbrace{\mathrm{SiLU}(\mathbf{x})\,\mathbf{W}_b^T}_{\text{residual}} + \underbrace{\sum_{d=0}^{D} \mathbf{C}_d \cdot T_d(\tanh(\mathbf{x}))}_{\text{Chebyshev spline}} \tag{7}$$

with recurrence $T_0 = 1$, $T_1 = x$, $T_{d+1} = 2xT_d - T_{d-1}$, and tanh normalizing inputs to $[-1, 1]$. Unlike standard MLP layers, which rely on fixed activation functions to approximate nonlinear mappings, KAN layers explicitly construct a functional basis through learnable polynomial expansions. This enables each connection to adaptively represent localized functional variations, effectively acting as a data-driven spectral decomposition mechanism.

In the context of structural dynamics, where system responses inherently contain multi-frequency components arising from modal interactions, this polynomial representation can be interpreted as an implicit basis expansion over frequency space. Consequently, KAN layers provide a natural mechanism to alleviate spectral bias even in the absence of explicit Fourier encoding. However, this increased representational flexibility comes at the cost of more complex gradient pathways, particularly when higher-order derivatives are involved in physics-informed loss functions. This trade-off between expressivity and gradient stability is a key aspect investigated in this study.

3.3 Model architectures with parameter matching

To ensure a rigorous and unbiased comparison, the model architectures are designed within a controlled experimental framework that isolates the effects of input representation and backbone structure. In particular, two orthogonal design factors are considered: (i) the use of Fourier feature encoding versus normalized time input, and (ii) the choice of backbone architecture (MLP versus KAN). A 2×2 factorial configuration refers to an experimental design in which two independent factors?input representation (Fourier vs. normalized time) and backbone architecture (MLP vs. KAN)?are systematically combined, resulting in four distinct model configurations that allow isolating the individual and combined effects of each factor. All models are constrained to ∼16,400 parameters (Table 3).

Remark 2 (Time normalization) Plain MLP and KAN-only receive $t/t_{\max} \in [0, 1]$ to prevent tanh saturation. Without this, $\tanh(60) \approx 1$ causes vanishing gradients. Fourier models are immune since sin / cos remain bounded. Omitting this normalization unfairly advantages Fourier models.

Table 3 Architecture specifications.

Model	Input	Hidden	Activation	Params
FPIKAN	$\Phi(t) \in \mathbb{R}^{40}$	KAN[32,32]	SiLU+Cheb	16,352
Fourier-MLP	$\Phi(t) \in \mathbb{R}^{40}$	MLP[110,110]	Tanh	16,831
Plain MLP	$t/t_{\max}$	MLP[90,90,90]	Tanh	16,651
KAN-only	$t/t_{\max}$	KAN[56,40]	SiLU+Cheb	16,352

3.4 Loss function and training protocol

The loss function is designed to enforce both data fidelity and physical consistency while maintaining stable gradient flow during training. In physics-informed learning, an improper balance between these components can lead to either overfitting to data or failure to satisfy the governing equations. Therefore, each term in the loss function is carefully constructed to address a specific aspect of the solution behavior. The total loss is:

$$\mathcal{L} = \mathcal{L}_{\text{data}} + \lambda_{\text{IC}}\mathcal{L}_{\text{IC}} + \lambda_v\mathcal{L}_{\text{vel}} + \lambda_p\mathcal{L}_{\text{phys}} \tag{8}$$

This term ensures that the network reproduces the reference solution accurately. However, relying solely on data loss is insufficient for nonlinear dynamic systems, as it does not guarantee physically consistent behavior, particularly in regions where data is sparse or highly nonlinear. $\mathcal{L}_{\text{data}} = N^{-1}\sum_i(\hat{u}(t_i) - u_{\text{ref}}(t_i))^2$ is the MSE. $\mathcal{L}_{\text{IC}} = (\hat{u}(t_0) - u_{\text{ref}}(t_0))^2$ enforces the *actual* initial value (not hardcoded zero). $\mathcal{L}_{\text{vel}} = (\partial\hat{u}/\partial t|_{t=0})^2$ enforces $\dot{x}(0) = 0$ via autograd for second-order ODEs. $\mathcal{L}_{\text{phys}}$ is the ODE residual at 150 collocation points (e.g., for Duffing: $\ddot{\hat{u}} + \delta\dot{\hat{u}} + \alpha_d\hat{u} + \beta_d\hat{u}^3 - \gamma_d\cos(\omega_d t) = 0$). Enforcing the correct initial condition is essential for dynamic systems, as even small deviations at t = 0 can propagate and amplify over time. Unlike many PINN implementations that assume zero initial states, this formulation explicitly preserves the true physical initial conditions of the system. Weights: $\lambda_{\text{IC}} = \lambda_v = 5$, $\lambda_p = 0.1$. Physics loss and velocity IC are applied *only* to FPIKAN and Fourier-MLP (PINN models) for Duffing and Van der Pol. Plain MLP and KAN-only are data-driven baselines. Best checkpoint is selected on $\mathcal{L}$ (total, not data alone). Optimizer: Adam [24], $\eta_0 = 3 \times 10^{-3}$, cosine annealing to 10^{-6}, 5,000 epochs. The choice of loss weights reflects a trade-off between physical consistency and numerical stability. A relatively smaller weight is assigned to the physics term ($\lambda_p = 0.1$) to prevent gradient domination by higher-order derivatives, which can destabilize training, especially in nonlinear systems.

3.5 Gradient analysis of KAN layers

To understand why FPIKAN underperforms Fourier-MLP, consider the gradient flow through a KAN layer. For the Chebyshev path in Eq. (7):

$$\frac{\partial}{\partial x}\left[\sum_d c_d T_d(\tanh(x))\right] = (1 - \tanh^2(x))\sum_d c_d T_d'(\tanh(x)) \tag{9}$$

where $T_d'(\xi) = dU_{d-1}(\xi)$ involves Chebyshev polynomials of the second kind with $|U_d(\xi)| \leq d+1$ on $[-1,1]$.

When $\mathcal{L}_{\text{phys}}$ requires $\partial^2\hat{u}/\partial t^2$, the gradient must traverse two nested chain-rule applications through KAN layers, producing terms involving $T_d''(\tanh(x))$. These second derivatives oscillate with amplitude proportional to $d^2(d^2-1)/3$ at domain boundaries [25], creating highly oscillatory gradient signals that compete with data-driven gradients. In MLP layers, $\partial^2 \tanh(x)/\partial x^2 = -2\tanh(x)(1-\tanh^2(x))$ is bounded and smooth, yielding well-conditioned gradient flow. This asymmetry explains the observed FPIKAN degradation under physics-informed training.

4 Results and Discussion

This section presents a comprehensive evaluation of the proposed architectures across all benchmark problems, with the aim of identifying the dominant factors governing predictive performance. The analysis is structured to progressively examine (i) overall accuracy, (ii) time-domain behavior, and (iii) frequency-domain fidelity, providing a multi-perspective assessment of model capabilities.

Particular emphasis is placed on understanding whether improvements arise from enhanced input representation or from architectural complexity. By leveraging the 2×2 factorial design introduced in Section 3, the results enable a controlled interpretation of how Fourier feature encoding and backbone choice influence both accuracy and stability.

In addition to conventional error metrics, the evaluation incorporates spectral and convergence analyses to capture behaviors that are not observable through time-domain error alone. This is especially critical for nonlinear structural dynamics, where multi-frequency interactions play a central role.

4.1 Comprehensive comparison

Several key observations emerge from the comprehensive results presented in Table 4. First, Fourier-MLP consistently achieves the lowest error across all benchmarks, demonstrating that the inclusion of Fourier feature encoding effectively eliminates spectral bias regardless of the underlying dynamical regime. In particular, near-zero L2 errors are obtained for the Duffing and Van der Pol systems, indicating that the model accurately captures both nonlinear and multi-frequency behavior.

Second, the performance of Plain MLP and KAN-only models highlights the severity of spectral bias in multi-frequency structural dynamics. For the Bouc?Wen and 3-DOF building benchmarks, both models yield L2 errors close to 1.0, which corresponds to a near-zero prediction. This indicates a complete failure to learn the underlying dynamics, rather than a minor approximation error.

Third, KAN-only demonstrates a clear advantage over Plain MLP in nonlinear oscillatory systems such as Duffing and Van der Pol, reducing the error by up to $3.5\times$. This confirms that learnable Chebyshev activations provide an implicit mechanism for capturing higher-frequency components even without explicit Fourier encoding.

Finally, the comparison between FPIKAN and Fourier-MLP reveals that combining Fourier encoding with KAN does not yield additional benefits and can even

Table 4 Comprehensive results (∼16,400 params each). Best L_2 per benchmark in **bold**.

Benchmark	Model	Mode	ε_{L_2}	$\varepsilon_{\text{peak}}$	$\varepsilon_{\text{spec}}$
Bouc–Wen	FPIKAN	data	0.250	0.000	0.152
	Fourier-MLP	data	**0.014**	**0.000**	**0.005**
	Plain MLP	data	0.998	0.969	0.995
	KAN-only	data	0.998	0.966	0.995
Duffing	FPIKAN	PINN	0.224	0.157	0.132
	Fourier-MLP	PINN	**0.000**	**0.000**	**0.000**
	Plain MLP	data	0.139	0.041	0.121
	KAN-only	data	0.074	0.001	0.063
Van der Pol	FPIKAN	PINN	0.179	0.075	0.103
	Fourier-MLP	PINN	**0.000**	**0.000**	**0.000**
	Plain MLP	data	0.574	0.118	0.483
	KAN-only	data	0.164	0.086	0.136
3-DOF Bldg.	FPIKAN	data	0.094	0.004	0.028
	Fourier-MLP	data	**0.003**	**0.003**	**0.000**
	Plain MLP	data	1.000	0.997	1.000
	KAN-only	data	1.000	0.994	1.000

degrade performance. Despite receiving identical Fourier inputs, FPIKAN consistently underperforms Fourier-MLP, suggesting that increased architectural complexity does not necessarily translate into improved accuracy when spectral bias is already mitigated at the input level.

Figure 1 provides a global visualization of model performance across all benchmark?architecture combinations. The dominance of Fourier-MLP is immediately evident, as indicated by the consistently dark green column, reflecting near-zero error across all problems. This uniform performance suggests that Fourier feature encoding effectively resolves the multi-frequency nature of structural response signals.

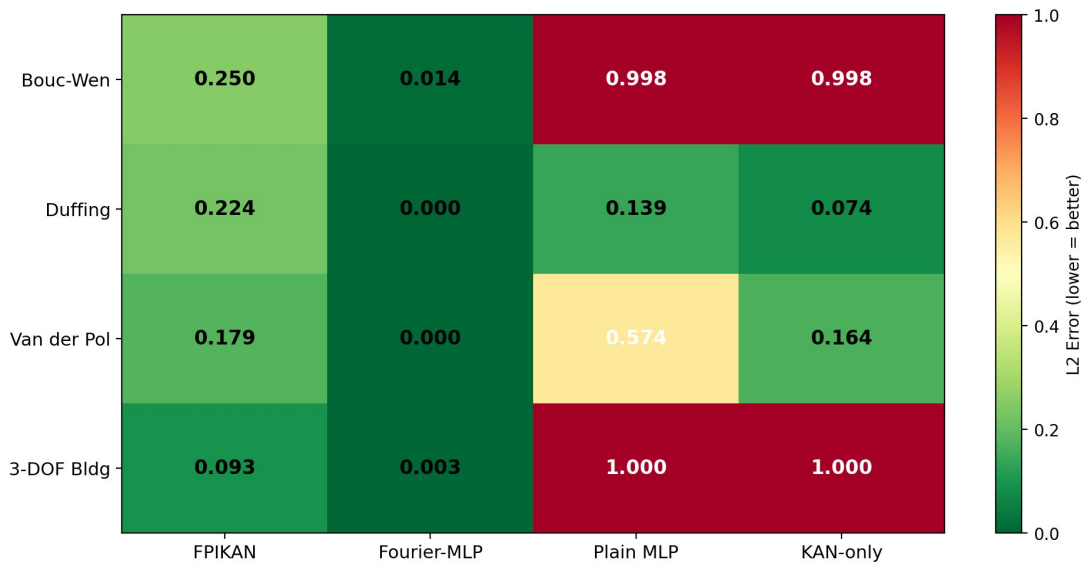


Fig. 1 L_2 error heatmap (≈16,400 params each). Green: low error; red: high error.

In contrast, Plain MLP and KAN-only models exhibit predominantly red regions for the Bouc?Wen and 3-DOF benchmarks, confirming their inability to capture high-frequency components. The near-uniform high error ($L_2 \approx 1.0$) indicates not partial inaccuracy but complete failure to learn the underlying dynamics.

4.2 Time-domain displacement response

Figure 2 shows displacement time histories. In the Bouc–Wen benchmark, Fourier-MLP closely tracks the hysteretic response while Plain MLP and KAN-only produce near-zero predictions ($\varepsilon_{L_2} \approx 1$). For Duffing, KAN-only captures the quasi-periodic pattern well even without Fourier features, but FPIKAN shows visible phase drift despite identical Fourier encoding. The 3-DOF building exhibits the highest frequency content, and non-Fourier model failure is most pronounced here.

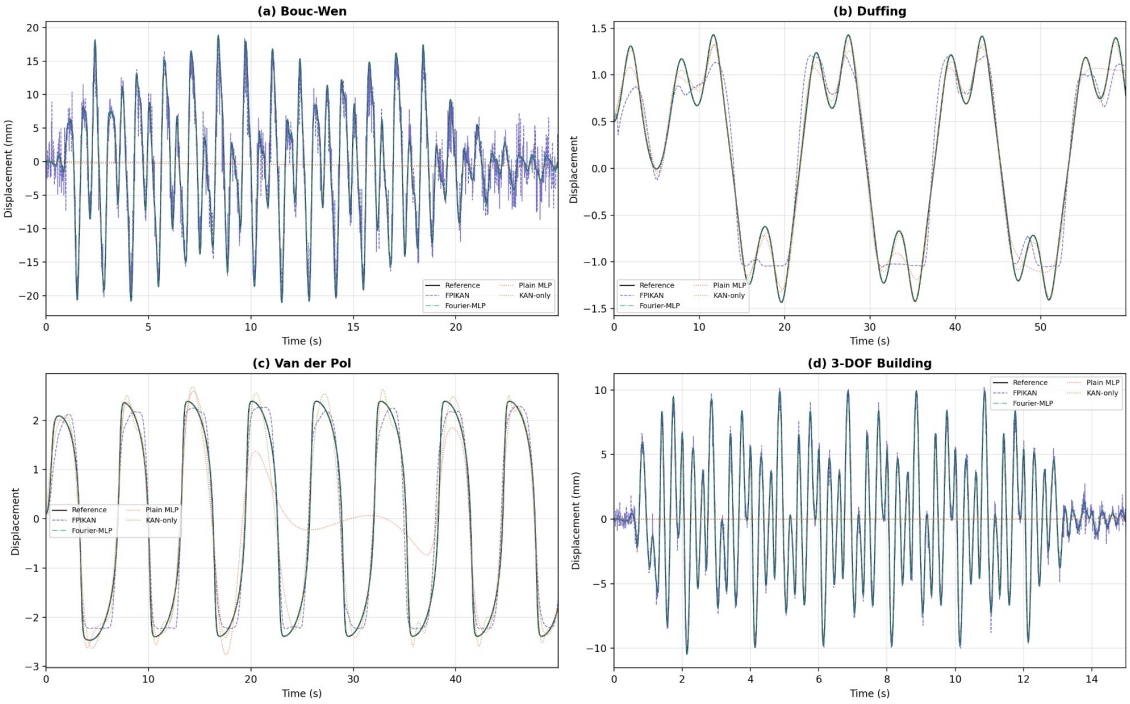


Fig. 2 Displacement response. (a) Bouc–Wen, (b) Duffing, (c) Van der Pol, (d) 3-DOF building.

The time-domain responses further highlight the fundamental differences between the models. Fourier-MLP accurately tracks both amplitude and phase across all benchmarks, demonstrating its ability to capture multi-frequency interactions inherent in nonlinear structural dynamics. In contrast, Plain MLP and KAN-only exhibit severe deviations in the Bouc-Wen and 3-DOF cases, where the response contains rich modal superposition. Their predictions collapse toward low-frequency or near-zero solutions, indicating an inability to represent high-frequency components. Interestingly, KAN-only shows reasonable performance in the Duffing system, where the dominant dynamics are less broadband. This suggests that the implicit spectral decomposition capability of Chebyshev activations can partially compensate for the lack of explicit Fourier encoding in simpler frequency regimes. However, FPIKAN introduces noticeable phase errors despite receiving Fourier features, reinforcing that increased architectural complexity does not guarantee improved temporal accuracy when gradient stability is compromised.

4.3 Point-wise error analysis

Figure 3 shows point-wise absolute error. Fourier-MLP maintains near-zero error throughout all benchmarks. For Bouc–Wen and 3-DOF, Plain MLP and KAN-only errors are as large as the signal itself, confirming total failure.

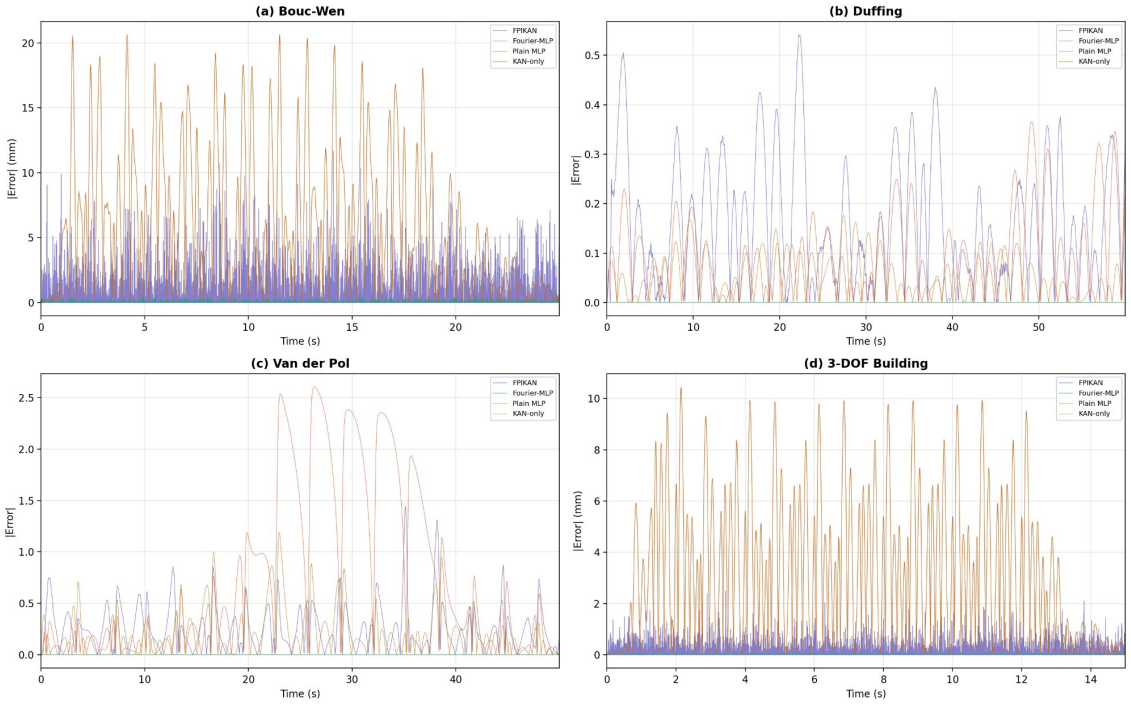


Fig. 3 Point-wise absolute error. (a) Bouc–Wen, (b) Duffing, (c) Van der Pol, (d) 3-DOF.

The point-wise error distribution reveals not only the magnitude but also the temporal structure of prediction errors. Fourier-MLP maintains consistently low error across the entire time domain, indicating stable learning of both transient and steady-state dynamics. For non-Fourier models, the error is uniformly high rather than localized, particularly in the Bouc?Wen and 3-DOF benchmarks. This confirms that the issue is not isolated misprediction but a global failure to capture the system dynamics. The absence of localized error peaks further suggests that these models are not approximating the solution with phase shifts or amplitude distortions, but instead failing to learn the underlying functional form altogether.

4.4 Training convergence

Figure 4 shows training MSE versus epoch. Fourier-MLP reaches MSE $< 10^{-7}$ in most benchmarks. FPIKAN converges initially but plateaus at 10^{-3}–10^{-5}, consistent with the gradient conditioning hypothesis (Section 3.5). Plain MLP and KAN-only plateau at MSE $\sim 10^{-1}$ for multi-frequency problems (Bouc–Wen, 3-DOF).

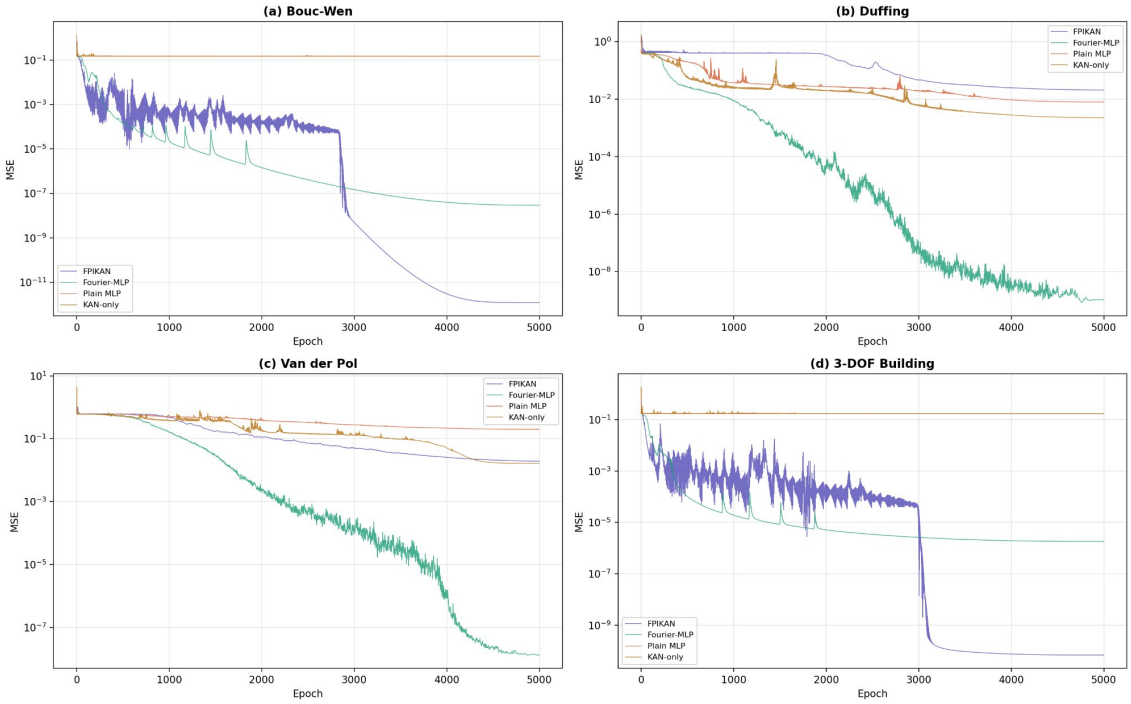


Fig. 4 Training convergence (MSE vs. epoch, log scale).

The convergence behavior provides additional insight into the optimization dynamics of each architecture. Fourier-MLP exhibits smooth and monotonic convergence, reaching very low MSE values across all benchmarks, which indicates well-conditioned gradient flow. FPIKAN demonstrates early convergence followed by stagnation at higher error levels. This behavior is consistent with the proposed gradient instability mechanism, where oscillatory gradients hinder effective optimization. Plain MLP and KAN-only plateau at relatively high error levels in multi-frequency benchmarks, confirming that spectral bias prevents further improvement even with prolonged training. These observations indicate that convergence characteristics are strongly influenced not only by model capacity but also by the interaction between input representation and gradient structure.

4.5 Spectral fidelity analysis

The Fourier spectra (Figure 5) reveal that Fourier-MLP reproduces the full reference spectrum across all frequency ranges. Non-Fourier models exhibit severe high-frequency attenuation—their spectra fall orders of magnitude below the reference above 2–3 Hz, directly visualizing spectral bias. Figure 6 summarizes errors in bar chart format.

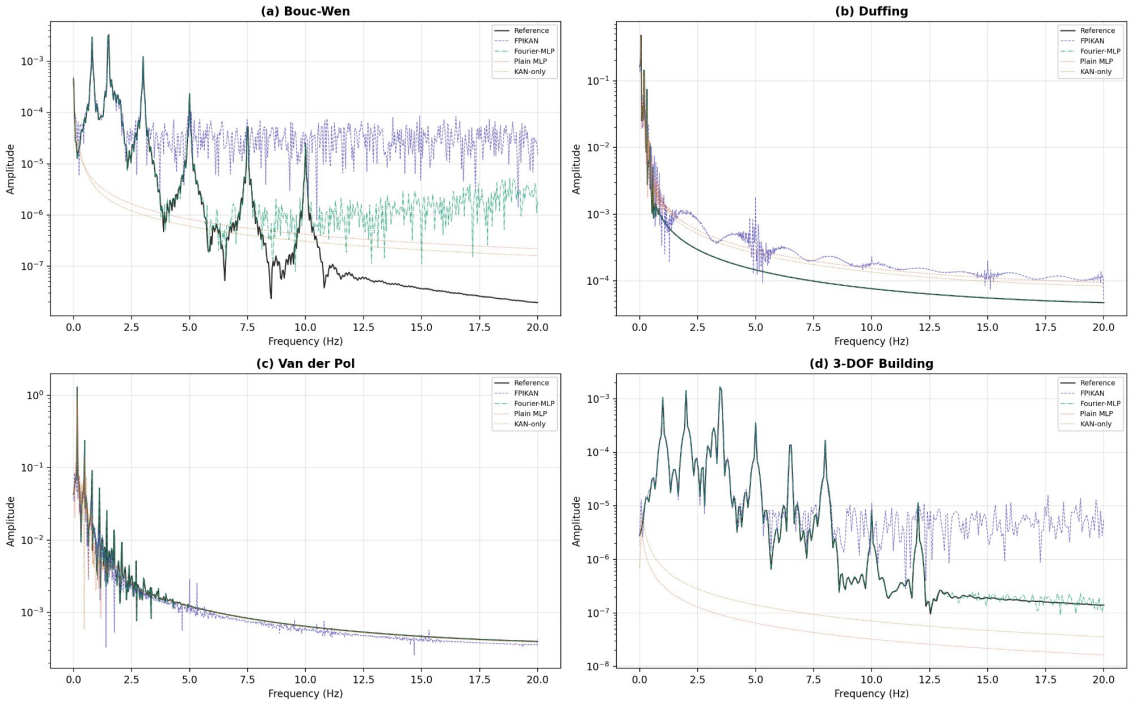


Fig. 5 Fourier amplitude spectra showing spectral bias in non-Fourier models.

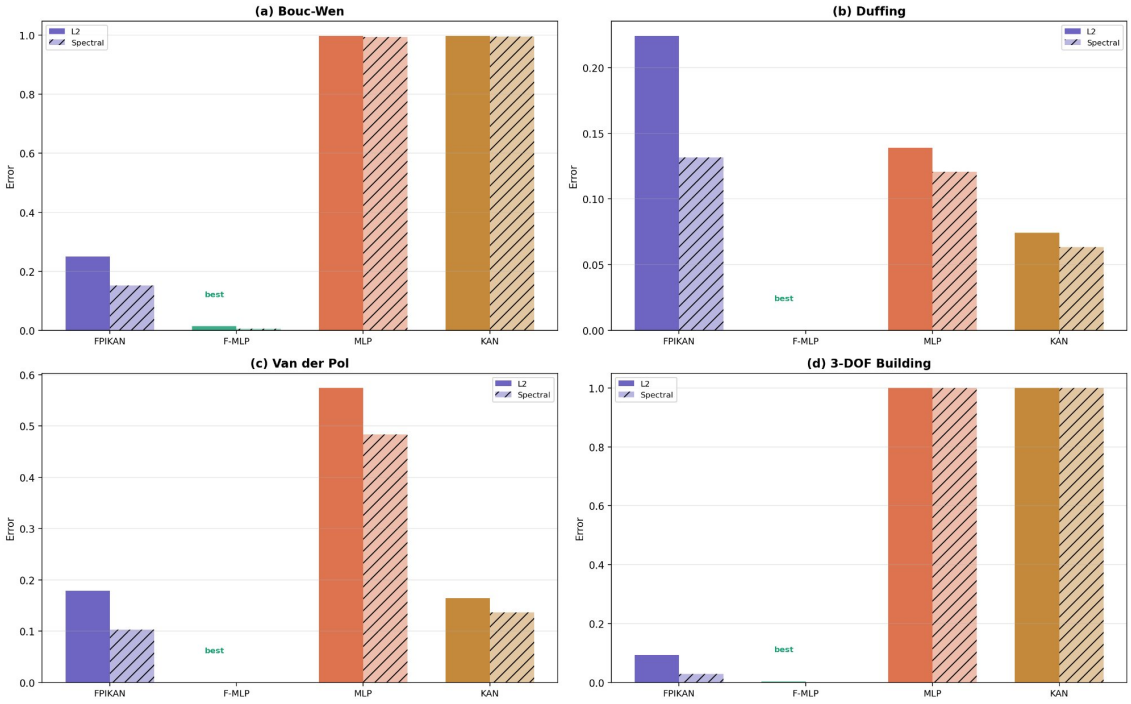


Fig. 6 L_2 (solid) and spectral error (hatched) comparison.

The spectral analysis provides direct evidence of how each model represents frequency content. Fourier-MLP accurately reconstructs the full frequency spectrum, including both dominant and higher-order harmonics, demonstrating that Fourier encoding effectively resolves spectral bias. Non-Fourier models exhibit severe attenuation of higher-frequency components, with their spectra dominated by a single low-frequency mode. This confirms that these models approximate the response as a smooth function, rather than capturing the true multi-frequency structure. FPIKAN partially recovers the spectral content but still shows reduced fidelity compared to Fourier-MLP, indicating that the backbone architecture limits the effective utilization of Fourier features. These results reinforce that spectral accuracy cannot be reliably inferred from time-domain metrics alone and must be explicitly evaluated in the frequency domain.

4.6 Ablation study

The 2×2 factorial design isolates component contributions (Table 5). The ablation results provide strong evidence that input representation dominates model performance. The dramatic reduction in error when adding Fourier features to MLP confirms that spectral bias is primarily an input-level limitation rather than an architectural one.

Furthermore, the degradation observed when combining Fourier encoding with KAN highlights a non-trivial interaction between representation and architecture. This suggests that certain architectural components may become redundant or even detrimental once the input space is sufficiently enriched. The ablation study demonstrates that increasing model complexity cannot compensate for inadequate input representation, and in some cases may introduce additional optimization challenges.

Table 5 Ablation: Fourier encoding effect and KAN vs. MLP comparison.

Benchmark	Fourier→MLP reduction	Fourier→KAN reduction	KAN vs MLP (w/ Fourier)	KAN vs MLP (no Fourier)
Bouc–Wen	+98.6%	+74.9%	MLP wins	≈ equal
Duffing	+100%	−201.7%	MLP wins	KAN 1.9×
Van der Pol	+100%	−9.0%	MLP wins	KAN 3.5×
3-DOF Bldg.	+99.7%	+90.7%	MLP wins	≈ equal

Finding 1: Fourier encoding is dominant. Adding Fourier features to MLP reduces ε_{L_2} by 98.6–100% across all benchmarks. This establishes that input representation, not network architecture, is the critical design choice.

Finding 2: KAN outperforms MLP without Fourier. KAN-only achieves lower ε_{L_2} than Plain MLP in Duffing (0.074 vs. 0.139, 1.9×) and Van der Pol (0.164 vs. 0.574, 3.5×). KAN's learnable Chebyshev activations function as implicit spectral decomposers, partially mitigating spectral bias without explicit Fourier features.

Finding 3: KAN with Fourier underperforms MLP with Fourier—and Fourier can even *hurt* KAN. Despite identical Fourier input, FPIKAN yields higher error than Fourier-MLP in all benchmarks. More strikingly, for Duffing, adding Fourier encoding to KAN *increases* the error from 0.074 (KAN-only) to 0.224 (FPIKAN)—a 3× degradation (−201.7% in Table 5). For Van der Pol, Fourier likewise slightly degrades KAN (0.164 → 0.179, −9.0%). This paradoxical result—where adding more information harms performance—points to a fundamental gradient incompatibility between Chebyshev polynomial layers and Fourier-encoded inputs under physics-informed training (Section 3.5).

5 Conclusions

This study systematically investigated the role of input representation and network architecture in physics-informed neural surrogates for nonlinear structural dynamics. Through a controlled experimental framework and comprehensive analysis, the results provide clear insight into the mechanisms governing model performance. Through parameter-matched experiments across four canonical benchmarks, this study establishes: Fourier feature encoding is the single most important architectural choice, reducing ε_{L_2} by 98.6–100% for MLP. Any PINN framework for structural dynamics should incorporate Fourier features as default. Fourier-MLP is the recommended architecture, achieving the lowest error in 4/4 benchmarks with ∼2× faster training than KAN alternatives and simple implementation. KAN provides value without Fourier

encoding, outperforming MLP by up to 3.5× through implicit spectral decomposition via learnable activations. This makes standalone KAN viable when frequency ranges are unknown a priori. Combining Fourier with KAN (FPIKAN) is counterproductive in the current Chebyshev-basis form. In the Duffing benchmark, adding Fourier features to KAN *triples* the L_2 error ($0.074 \rightarrow 0.224$, -201.7%). The oscillatory gradient pathways of high-degree Chebyshev polynomials conflict with second-order physics residual terms. Future work should explore B-spline or wavelet-based KAN variants with bounded, localized derivatives. Methodological rigor is essential. Seven specific implementation bugs were identified in common PINN codes (hardcoded zero IC, missing velocity IC, unmatched parameters, missing time normalization, physics loss applied to all models, checkpoint on wrong loss, incorrect PINN labeling) that individually produce misleading results.

Future work should investigate B-spline [9] and wavelet-based KAN variants, extension to stochastic earthquake records (PEER NGA-West2), and multi-output formulations for hidden hysteretic variable prediction.

6 Declarations

6.1 Conflict of interest:

The authors have no conflicts of interest to declare that are relevant to the content of this article.

6.2 Funding:

No funding was received for conducting this study.

6.3 Author's contribution:

[Salih Berkan AYDEMIR]—Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing-Original Draft, Visualization.

6.4 Acknowledgement:

Not applicable.

6.5 Data availability statement:

All data supporting the findings are included within the article. Source code is available from the corresponding author upon reasonable request.

6.6 Usage of AI tools:

The authors used generative AI for code development assistance and manuscript drafting support. All scientific decisions, experimental design, and result interpretation were performed by the authors.

References

[1] Naeim, F. (2007). Dynamics of structures–theory and applications to earthquake engineering. Earthquake Spectra, 23(2), 491-492.

[2] Newmark, N. M. (1959). A method of computation for structural dynamics. Journal of the engineering mechanics division, 85(3), 67-94.

[3] Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422-440.

[4] Raissi, M., Perdikaris, P., & 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.

[5] Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y. & Courville, A. (2019, May). On the spectral bias of neural networks. In International conference on machine learning (pp. 5301-5310). PMLR.

[6] Xu, Z. Q. J., Zhang, Y., Luo, T., Xiao, Y., & Ma, Z. (2019). Frequency principle: Fourier analysis sheds light on deep neural networks. arXiv preprint arXiv:1901.06523.

[7] Tancik, M., Srinivasan, P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ravi Ramamoorthi, Jonathan Barron, & Ng, R. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. Advances in neural information processing systems, 33, 7537-7547.

[8] Wang, S., Wang, H., & Perdikaris, P. (2021). On the eigenvector bias of Fourier feature networks: From regression to solving multi-scale PDEs with physics-informed neural networks. Computer Methods in Applied Mechanics and Engineering, 384, 113938.

[9] Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Solja?i?, M., Marin Solja?i?, Thomas Y. Hou, & Tegmark, M. (2024). Kan: Kolmogorov-arnold networks. arXiv preprint arXiv:2404.19756.

[10] Shukla, K., Toscano, J. D., Wang, Z., Zou, Z., & Karniadakis, G. E. (2024). A comprehensive and FAIR comparison between MLP and KAN representations for differential equations and operator networks. Computer Methods in Applied Mechanics and Engineering, 431, 117290.

[11] Rao, X., Liu, Y., He, X., & Hoteit, H. (2025). Physics-informed Kolmogorov-Arnold networks to model flow in heterogeneous porous media with a mixed pressure-velocity formulation. Physics of Fluids, 37(7).

[12] Tao, D., Li, J., Du, K., & Lu, Z. (2026). A fourier-enhanced physics-informed Kolmogorov?Arnold network for multi-frequency seismic response analysis of structures. Bulletin of Earthquake Engineering, 1-23.

[13] Wen, Y. K. (1976). Method for random vibration of hysteretic systems. Journal of the engineering mechanics division, 102(2), 249-263.

[14] Baber, T. T., & Noori, M. N. (1985). Random vibration of degrading, pinching systems. Journal of Engineering Mechanics, 111(8), 1010-1026.

[15] Foliente, G. C. (1995). Hysteresis modeling of wood joints and structural systems. Journal of Structural Engineering, 121(6), 1013-1022.

[16] Ismail, M., Ikhouane, F., & Rodellar, J. (2009). The hysteresis Bouc-Wen model, a survey. Archives of computational methods in engineering, 16(2), 161-188.

[17] Duffing, G. (1918). Erzwungene Schwingungen bei veranderlicher Eigenfrequenz und ihre technische Bedeutung (No. 41-42). Vieweg.

[18] Ueda, Y. (1979). Randomly transitional phenomena in the system governed by Duffing's equation. Journal of Statistical Physics, 20(2), 181-196.

[19] Moon, F. C., & Holmes, P. J. (1979). A magnetoelastic strange attractor. Journal of Sound and Vibration, 65(2), 275-296.

[20] Lifshitz, R., & Cross, M. C. (2008). Nonlinear dynamics of nanomechanical and micromechanical resonators. Reviews of nonlinear dynamics and complexity, 1(1), 1-52.

[21] Van der Pol, B. (1926). LXXXVIII. On ?relaxation-oscillations?. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11), 978-992.

[22] Strogatz, S. H. (2024). Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. Chapman and Hall/CRC.

[23] Period-doubling cascades and devil's staircases of the driven Van der Pol oscillator, Physical Review A,

[24] Diederik, P. K., & Ba, J. L. (2015). Adam: a Method for Stochastic Optimization International Conference On Learning Representations, 2015.

[25] Mason, J. C., & Handscomb, D. C. (2002). Chebyshev polynomials. Chapman and Hall/CRC.