Interpretable Deep Learning for Detecting and Quantifying Nonlinear Mode Interactions from Time-Series Response Data

Bayan Abusalameh [1] **Gianmarco Mengaldo** [1]

[1] *College of Design and Engineering, National University of Singapore, Singapore 119077*. Correspondence to: Gianmarco Mengaldo mpegim@nus.edu.sg.

1. Introduction

Reliably detecting, characterizing, and quantifying nonlinear mode interactions in vibrating structures remains a central challenge in computational mechanics, especially when only response time series are available and interaction signatures may be subtle, transient, or noise-corrupted. We propose a workflow that couples supervised deep learning with post-hoc interpretability to infer interaction presence and interaction-regime strength directly from raw multichannel response histories, while exposing the time-localized evidence that drives each decision.

The approach builds on our prior work on interpretable deep learning for structural nonlinearities in single-degree-of-freedom settings [1] and extends it to multi-modal interaction regimes. Training data are generated by solving parameterized equations of motion for multi-mode systems while systematically varying detuning, damping, coupling strength, forcing type, amplitude, and measurement noise. Labels correspond to interaction regimes (e.g., non-interacting, weakly interacting, strongly interacting), enabling learning without handcrafted diagnostics.

2. Method

For classification, we evaluate multiple time-series model families: Convolutional Neural Networks (CNNs), Temporal Convolutional Networks (TCNs) [2], and Bidirectional Long Short-Term Memory networks (BiLSTMs) [3, 4]. CNN-style backbones provide robust local feature extraction under consistent training protocols; BiLSTMs serve as a complementary recurrent baseline to test whether conclusions depend on a single model class; and TCNs provide a convolutional sequence alternative with long effective receptive fields.

To connect predictions to physics, we apply post-hoc interpretability to produce relevance maps over the input time series, highlighting temporal segments (and associated response components) most influential for class decisions. We use attribution methods such as Integrated Gradients [5] and gradient-based localization approaches (e.g., Grad-CAM-style methods) [6] to support auditable explanations aligned with interaction mechanisms, including intervals suggestive of energy exchange, modulation products, and shifts in dominant response content.

3. Results and significance

Across a wide sweep of forcing conditions and signal-to-noise ratios, the workflow detects and categorizes nonlinear mode-interaction regimes from raw time series while producing explanations that remain interpretable and engineering-relevant. By combining classification with auditable evidence, the approach supports trustworthy AI-assisted diagnostics for nonlinear dynamics, where adoption depends on traceability and physical plausibility—not only predictive accuracy.

Acknowledgments

References

[1] Bayan Abusalameh, Gianmarco Mengaldo, and Jiawen Wei. Data-driven classification of structural nonlinearities using interpretable deep learning on time series, 2025. Available at SSRN 5717653.

[2] Colin Lea, René Vidal, Austin Reiter, and Gregory D. Hager. Temporal convolutional networks for action segmentation and detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[3] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

[4] Mike Schuster and Kuldip K. Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.

[5] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2017.

[6] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.