https://lib.uliege.be https://matheo.uliege.be

Master thesis and internship[BR]- Master's thesis : Machine learning for experimental bifurcation analysis[BR]- Integration internship

Auteur : Bourdouch, Grégoire
Promoteur(s) : Kerschen, Gaëtan; Geurts, Pierre
Faculté : Faculté des Sciences appliquées
Diplôme : Master en ingénieur civil en aérospatiale, à finalité spécialisée en "aerospace engineering"
Année académique : 2024-2025
URI/URL : http://hdl.handle.net/2268.2/23382

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

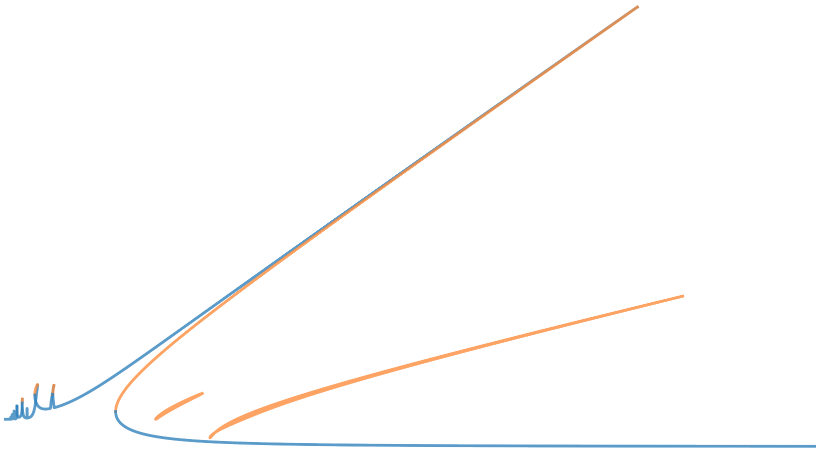
Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

University of Liège - Faculty of Applied Sciences

Master's thesis

for the degree of Master of Science in Aerospace Engineering

Machine learning for experimental bifurcation analysis



Author

Grégoire Bourdouch

Supervisors

Prof. Gaëtan Kerschen
University of Liège

Prof. Pierre Geurts
University of Liège

Academic Year 2024–2025

Jury

President

Prof. Gaëtan Kerschen
University of Liège

Supervisors

Prof. Gaëtan Kerschen
University of Liège

Prof. Pierre Geurts
University of Liège

Member

Prof. Loïc Salles
University of Liège

June 2025

LIÈGE université
Sciences Appliquées

Abstract

Understanding and predicting the dynamic vibrational response of mechanical systems remains a core challenge in structural dynamics. Real-world structures frequently exhibit nonlinear behaviour, resulting in complex phenomena such as primary and secondary resonances, multistability, and the emergence of unstable response branches. These features lie beyond the reach of standard experimental modal analysis (EMA), which, based on linear assumptions and typically relying on swept-sine excitation, can only capture the stable portions of the frequency response. As a result, crucial information about bifurcations and unstable dynamics remains experimentally inaccessible. Although control-based continuation methods have been developed to overcome this limitation, their practical application is hindered by the need for precise controller tuning and complex feedback architectures.

This thesis proposes a novel data-driven methodology that bridges this gap by combining standard EMA techniques with deep learning. Specifically, it investigates whether unstable branches of nonlinear frequency response curves (FRCs) can be predicted solely from experimental swept-sine measurements, which contain only the stable portions of the response. To the best of the author's knowledge, this is the first approach that enables the prediction of unstable branches of nonlinear frequency response curves from experimental swept-sine data alone.

To this end, a supervised learning framework is developed and applied to the Duffing oscillator, a one-degree-of-freedom nonlinear system with cubic stiffness. A convolutional neural network is trained on synthetic datasets, with input data generated from binary image representations of swept-sine envelopes. The model successfully predicts not only the fundamental resonance but also subharmonic and superharmonic resonances, effectively reconstructing the full FRC.

Crucially, the methodology is validated on an experimental electronic Duffing oscillator, confirming that unstable branches can indeed be predicted from swept-sine tests alone. The approach is then extended to other oscillator types and nonlinearity forms, demonstrating promising generalization capabilities.

Overall, this work presents a practical and effective alternative to control-based methods for the analysis of nonlinear experimental data. It demonstrates the feasibility of integrating standard EMA techniques with machine learning to predict otherwise inaccessible dynamic features.

Keywords: Experimental modal analysis, Machine learning, Nonlinear dynamics, Frequency response curve, Bifurcation, Secondary resonances, Isolated responses

Résumé

Comprendre et prédire la réponse vibratoire dynamique des systèmes mécaniques demeure un enjeu central en dynamique des structures. Contrairement aux systèmes linéaires, les structures réelles présentent souvent un comportement non linéaire, donnant lieu à des phénomènes complexes tels que les résonances primaires et secondaires, le multi-stabilité, ainsi que l'apparition de branches de réponse instables. Ces caractéristiques échappent aux méthodes classiques d'analyse modale expérimentale, qui, fondées sur des hypothèses linéaires et reposant généralement sur des essais de type balayage sinusoïdal, ne permettent d'identifier que les solutions stables de la courbe de réponse fréquentielle. Par conséquent, des informations cruciales concernant les bifurcations et les dynamiques instables restent inaccessibles expérimentalement. Bien que des méthodes de continuation expérimentale aient été développées pour accéder à ces régions, leur mise en œuvre reste limitée en raison de la complexité de la conception des contrôleurs et du réglage précis des gains.

Ce mémoire propose une méthodologie innovante, fondée sur l'apprentissage automatique, qui vise à combler cette lacune en exploitant des données issues de tests expérimentaux classiques. Plus précisément, l'approche vise à prédire les branches instables des courbes de réponse fréquentielle non linéaires à partir de mesures obtenues par balayage sinusoïdal. À la connaissance de l'auteur, il s'agit de la première méthode permettant de prédire les branches instables d'une courbe de réponse fréquentielle non linéaire à partir de ce seul type de mesure expérimentale.

Pour ce faire, une méthode d'apprentissage supervisé est développée et appliquée à l'oscillateur de Duffing, un système non linéaire à un degré de liberté caractérisé par une raideur cubique. Un réseau de neurones convolutifs est entraîné sur des données synthétiques, construites à partir d'images binaires représentant les enveloppes des réponses mesurées issues de balayages sinusoïdaux. Le modèle permet de prédire avec succès non seulement la résonance fondamentale, mais également les résonances secondaires comme les super-harmoniques et sous-harmoniques, reconstituant ainsi l'ensemble de la courbe de réponse fréquentielle.

La méthodologie est ensuite validée expérimentalement sur un oscillateur électronique reproduisant le comportement de l'oscillateur de Duffing, confirmant la capacité du modèle à reconstruire l'ensemble de la courbe de réponse fréquentielle, y compris ses branches instables, à partir de données expérimentales. Enfin, l'approche est étendue à d'autres types d'oscillateurs et de non-linéarités, mettant en évidence un potentiel de généralisation prometteur.

Ce travail constitue une alternative pratique et prometteuse aux méthodes de continuation expérimentale pour l'analyse des systèmes non linéaires. Il démontre la faisabilité d'intégrer des techniques classiques, telles que le balayage sinusoïdal, avec des approches d'apprentissage automatique, afin de prédire des dynamiques inaccessibles sans recours à des dispositifs de contrôle actif.

Mots-clés : Analyse modale expérimentale, Apprentissage automatique, Dynamique non linéaire, Réponse fréquentielle, Bifurcation, Résonances secondaires, Isolats

Acknowledgment

I would like to express my sincere gratitude to my supervisor, Prof. Gäetan Kerschen, for his insightful guidance, constant encouragement, and unwavering support throughout the course of this research. His expertise in nonlinear dynamics and his mentorship have been invaluable to the successful completion of this work.

I am also profoundly thankful to Prof. Pierre Geurts and Prof. Loïc Salles for serving on the jury and for their constructive feedback. My appreciation also extends to all the members of the S3L lab for their support and for sharing their knowledge, which helped create a welcoming and stimulating working environment.

Finally, I would like to thank Camille, my family, and my friends for their ongoing encouragement, love, and support throughout my studies.

Contents

List of symbols and acronyms

Symbols

c	Damping coefficient
F	Forcing amplitude
$\bar{F}$	Nondimensionalised forcing amplitude
$\tilde{F}$	Mass-normalised excitation amplitude
k_1	Linear stiffness coefficient
k_2	Quadratic stiffness coefficient
k_3	Cubic stiffness coefficient
m	Mass
N_H	Harmonic truncation order
t	Time
y	Displacement
$\dot{y}$	Velocity
$\ddot{y}$	Acceleration
y_0, $\dot{y}_0$	Initial displacement and velocity
α	Focal loss coefficient
α_2	Mass-normalised quadratic stiffness coefficient
α_3	Mass-normalised cubic stiffness coefficient
β	Non-dimensional quadratic stiffness parameter
γ	Focal loss class-balancing parameter
Ω	Forcing frequency
ω_0	Natural frequency
ζ	Damping ratio
τ	Normalised time / Decision threshold
ϕ	Phase angle
σ	Normalisation scaling factor
$\bar{y}, \bar{\omega}$	Nondimensional displacement and frequency
$\mathcal{L}$	Loss function

Acronyms

(A)CBC	(Arclength) Control-Based Continuation
BCE	Binary Cross-Entropy
CNN	Convolutional Neural Network
DL	Deep Learning
(S)DOF	(Single-)Degree-of-Freedom
FRC	Frequency Response Curve
HB	Harmonic Balance
ML	Machine Learning
PLL	Phase-Locked Loop
RCT	Response-Controlled Stepped-Sine Testing
ReLU	Rectified Linear Unit

CHAPTER 1

Introduction

1.1 Context

Understanding and predicting the dynamic behaviour of mechanical systems is a fundamental task in engineering. Whether for designing new systems or analysing existing ones, the objective is to develop mathematical models capable of accurately reproducing a system's response to various types of excitations. These models are crucial for simulating the behaviour of conceptual systems that have yet to be physically realised and for predicting the response of known structures under untested loading scenarios.

For systems exhibiting linear behaviour, the superposition principle provides a powerful simplification: the response to a combined excitation can be represented as the sum of the responses to its individual components. This property enables the use of frequency domain techniques, such as stepped or swept sine testing and broadband excitation, to identify the system's frequency response function. These methodologies form the basis of linear modal analysis, a field supported by well-established theoretical, numerical, and experimental frameworks [23, 24, 57].

However, linear approximations fall short when addressing the complexities of real-world structures, which are often inherently nonlinear. Nonlinearities may arise from large structural deformations, the use of advanced materials such as foams, composites, or elastomers, as well as other sources, including geometric effects, inertial couplings, contact interfaces, and non-classical damping mechanisms like friction or hysteresis. While the theoretical and numerical study of nonlinear dynamics has a long academic history, it is only in recent years that the industrial sector has begun to show significant interest in this field. This growing attention is driven mainly by the increasing demand for lightweight and high-performance structures, which operate under conditions where nonlinear effects become prominent and can no longer be neglected. As a result, there is a clear shift toward incorporating nonlinear analysis into the design and diagnostic processes of complex engineering systems.

In cases of weak nonlinearity, standard linear identification methods may still provide useful insights. Yet, for many practical systems, these techniques fail to capture the full spectrum of dynamic responses, particularly when subjected to multi-harmonic or high-amplitude excitations. The absence of a superposition principle in nonlinear systems significantly complicates the process of response characterisation.

To manage this complexity, it is common to focus on specific classes of responses. One widely studied class is the family of frequency response curves (FRCs), which represent the steady-state system responses at a fixed excitation amplitude across a range of frequencies. These curves serve as the nonlinear counterpart to the frequency response function (FRF) in linear systems [51]. Another type of response branch is the backbone curves, which are a collection of responses at resonance, showing the frequency-amplitude relation of nonlinear normal modes [77].

Nonlinear FRCs exhibit rich and intricate phenomena, including bifurcations and chaotic responses. Two notable features are of particular interest in this work:

- **Multistability and bifurcations:** A single excitation input can correspond to multiple possible steady-state responses. This is associated with bifurcations, which lead to phenomena such as jumps, where a small change in frequency or amplitude causes a sudden transition between response branches [51].

- **Unstable responses:** Some system responses are inherently unstable and cannot be observed directly in standard open-loop experiments due to their sensitivity to initial conditions and external perturbations. However, such responses can be accessed using control-based continuation methods, which stabilise and trace these otherwise inaccessible branches. Unstable solutions are essential for understanding the global dynamics of nonlinear systems, as they often serve as transitional pathways between stable branches and help reveal the structure of bifurcations in the response landscape [27].

Figure 1.1 illustrates the frequency response of a harmonically forced Duffing oscillator, a canonical single-degree-of-freedom system exhibiting cubic stiffness nonlinearity. The response is characterised by the coexistence of multiple solutions for certain frequency ranges, including both stable and unstable branches of solutions. The figure highlights the phenomena of jump discontinuities that occur during forward (sweep-up) and backward (sweep-down) frequency excitations due to bifurcations.

Additionally, the plot reveals the presence of complex nonlinear features such as superharmonic and subharmonic resonances. In particular, subharmonic resonances often manifest as isolated solution branches. These features are indicative of the rich dynamical behaviour introduced by the system's nonlinearity. This work focuses on exploring and predicting such nonlinear behaviours, with particular attention to the unstable regions of FRCs that are essential for a comprehensive understanding of the system but remain inaccessible from conventional experimental observation.

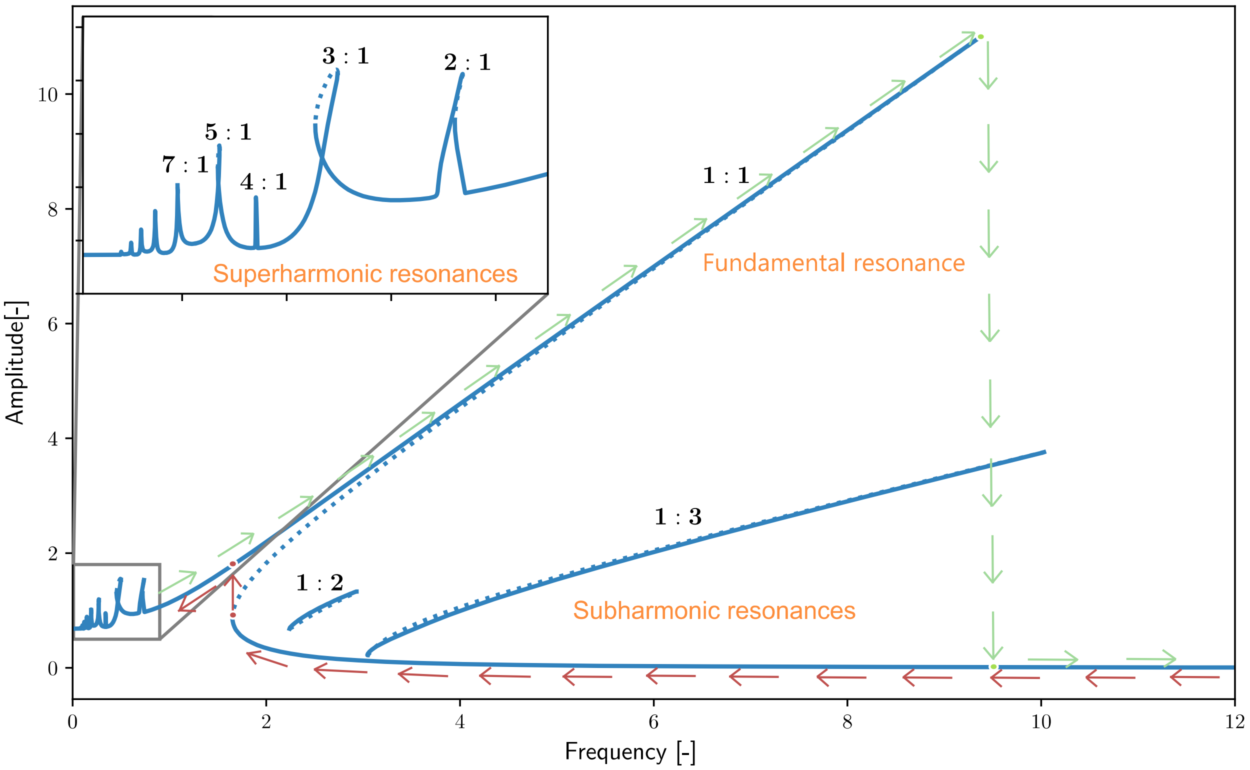


Figure 1.1: Frequency response curve of a Duffing oscillator governed by the equation $\ddot{y} + 0.01\dot{y} + y + y^3 = \sin(\omega t)$, computed using a harmonic balance-based numerical continuation method [18]. Solid lines denote stable branches, while dashed lines indicate unstable solutions. The arrows illustrate the paths corresponding to forward (—) and backward (—) frequency sweeps. Subharmonic resonances appear as detached solution branches.

1.2 Control-based continuation methods for nonlinear vibrations

Control-Based Continuation (CBC) is an experimental methodology designed to trace the full spectrum of nonlinear responses of a physical system, including unstable branches of responses. Initially proposed in theory by Sieber and Krauskopf [68] and later realised experimentally by Barton and Burrow [7], CBC embeds the system under test within a feedback loop. A proportional–integral–derivative (PID) controller adjusts the input to enforce a desired output signal, enabling the stabilisation of periodic responses that would otherwise be inaccessible. However, classical CBC implementations rely on finite-difference approximations of output sensitivities (i.e., Jacobians), which are prone to noise and can reduce the method's robustness in practical experiments.

To alleviate the computational and noise-related difficulties of Jacobian estimation, a simplified variant referred to as s-CBC was developed [7]. s-CBC limits the continuation to the fundamental harmonic of the system's response. Higher-order harmonics, which contribute to the system's full nonlinear behaviour, are copied directly from the output into the controller reference, ensuring non-invasiveness. This makes the method faster and more robust to noise, but at the cost of completeness: s-CBC does not capture secondary resonances or isolated branches, and post-processing is required to reconstruct the full nonlinear frequency response curve.

Building upon the idea of s-CBC, Abeloos proposed a more systematic and derivative-free framework called Arclength Control-Based Continuation (ACBC) [1]. ACBC replaces numerical differentiation with a secant-based predictor and employs a fixed-arclength constraint in the continuation procedure. Additionally, adaptive filtering is used to ensure the non-invasiveness of the control action. ACBC is capable of traversing bifurcations and identifying unstable branches in a robust and automated fashion. However, it remains limited to continuation in the fundamental harmonic and cannot directly target secondary resonances or disconnected solution branches (isolas).

This limitation was addressed in a subsequent extension known as extended ACBC (x-ACBC), developed by Spits [71]. x-ACBC generalises the continuation framework by allowing the use of non-fundamental harmonic components, such as superharmonics or subharmonics, as continuation parameters. This enables the direct experimental detection of complex features in the response landscape, including secondary resonances and isolas. x-ACBC thus represents a significant expansion in the capabilities of control-based continuation, opening access to a wider class of nonlinear phenomena.

Despite these advancements, all control-based continuation methods share a common set of challenges. Primary among them is the selection of controller gains. The PID gains must be carefully tuned to stabilise the desired orbits without making the control action invasive or destabilising the system through unintended feedback dynamics. This tuning process is largely heuristic and sensitive to the system's characteristics. Improper gain selection can lead to poor convergence, incorrect stabilisation, or even the failure to track unstable branches. Moreover, different harmonic components may require different gain settings, especially in x-ACBC, where each targeted resonance might exhibit unique stability properties. The design of adaptive, real-time gain-tuning strategies remains an open area of research critical to the broader adoption of these methods.

An alternative approach to nonlinear experimental continuation is the Phase-Locked Loop (PLL) method [70, 84]. Rather than adjusting the amplitude of the excitation, PLL testing fixes the phase lag between the excitation and the system response using a feedback controller. The excitation frequency is automatically adjusted until the desired phase lag is achieved, which allows the method to efficiently trace backbone curves and, in some cases, lock onto superharmonic resonances. PLL methods are especially attractive due to their speed and simplicity when exploring primary resonances.

However, PLL testing also suffers from notable limitations. Its performance is highly sensitive to phase-detection noise and the precision of the phase-estimation algorithm, especially far from resonance, where the phase slope becomes steep. Moreover, in multi-degree-of-freedom (MDOF) systems, different modes can share the same phase lag, leading to ambiguous or folded response surfaces and mode-jumping phenomena. The tuning of PLL controller gains (typically proportional–integral) is also critical: overly aggressive gains can lead to instability or oscillations in the loop, while conservative gains may result in slow convergence or failure to lock.

A final control method is the Response-Controlled Stepped-Sine Testing (RCT) developed by Karağaçlı and Özgüven [35]. This technique identifies nonlinear modal parameters by maintaining constant displacement amplitude during stepped-sine testing, yielding quasi-linear frequency response functions for conventional modal analysis. However, it requires well-separated modes and no internal resonances, limiting its use with strong modal coupling.

In summary, the CBC family of methods and PLL testing offer powerful tools for the experimental continuation of nonlinear structures. CBC, s-CBC, ACBC, and x-ACBC represent a progression toward more robust and comprehensive schemes, each addressing challenges related to noise sensitivity and stability. Nevertheless, all remain subject to the delicate trade-off between control robustness and non-invasiveness, primarily governed by gain tuning. PLL, while simpler and faster for backbone tracking, is constrained by phase-related limitations and lacks the generality needed to fully explore the nonlinear response landscape. Complementing these approaches, the RCT method offers a practical alternative by producing quasi-linear frequency response functions through displacement-controlled excitation, enabling the identification of nonlinear modal parameters using standard modal analysis tools. However, RCT is limited to systems with well-separated modes and no internal resonances. The development of robust, adaptive, and broadly applicable control strategies thus remains a vital direction for future work in experimental nonlinear dynamics.

Emergence of AI and motivation of this research

The rapid emergence and evolution of artificial intelligence (AI) have profoundly influenced engineering disciplines, motivating the exploration of data-driven approaches in traditionally model-based domains. While public attention has recently focused on large language models (LLMs) like ChatGPT, AI encompasses a broad spectrum of techniques, with machine learning (ML) and deep learning (DL) being the most widely adopted in industrial applications. These methods have demonstrated remarkable capabilities in tasks ranging from image recognition and natural language processing to anomaly detection and predictive maintenance.

In engineering, and particularly in the field of structural dynamics, ML and DL have started to complement and, in some cases, surpass classical modelling techniques [14]. Applications include real-time structural health monitoring [4], fault detection in rotating machinery [33], and many other applications. The transition from analytical or physics-based models to data-driven approaches is driven by the ability of AI systems to learn complex nonlinear mappings from data, handle noisy and incomplete measurements, and adapt to varying operating conditions without extensive manual tuning.

This shift is increasingly evident in nonlinear vibration analysis, where ML has been successfully used for modal identification [17, 64] and damage detection [34]. More recently, researchers have begun applying deep learning to capture nonlinear phenomena, including bifurcations and multistability [29], which are traditionally difficult to analyse using linear frameworks. Despite these promising results, the application of AI to nonlinear frequency response analysis, particularly in predicting unstable branches and isolated solutions, remains largely unexplored. While reinforcement learning (RL) has not yet been applied in the context of CBC or PLL techniques, its potential for enhancing these methods is promising. In particular, RL-based control strategies have been successfully employed to automatically tune the gains of PID controllers [11], enabling adaptive and robust control in nonlinear and uncertain environments. These capabilities are directly relevant to CBC and PLL. By exploiting RL to adapt control gains or policies in real-time, it may be possible to improve the efficiency, robustness, and automation of experimental continuation procedures. This emerging intersection of AI and control theory could potentially play a key role in advancing the experimental exploration of nonlinear dynamics.

Motivated by these developments, this research explores the potential of deep learning to support and enhance nonlinear vibration analysis, particularly in scenarios where standard methods face limitations. This work investigates how AI can serve as a complementary tool, offering new perspectives for identifying complex dynamic behaviours such as unstable responses.

1.3 Objectives

Building on the motivations outlined above, this research aims to develop a data-driven methodology for identifying the nonlinear frequency response of systems subjected to harmonic excitation. The approach leverages ML techniques and is validated through both numerical and experimental simulations.

This master's thesis seeks to address the following key questions:

1. Can machine learning approaches bridge the gap between standard experimental model analysis (EMA) testing and nonlinear frequency response?
2. How well does the machine learning model generalise when applied to both experimental data and nonlinear numerical systems not represented in the training set?

Figure 1.2 provides a high-level conceptual overview of the conducted research. It illustrates the general idea underpinning the approach: leveraging standard excitation data typically accessible through standard experimental techniques to predict complex nonlinear behaviours using machine learning.

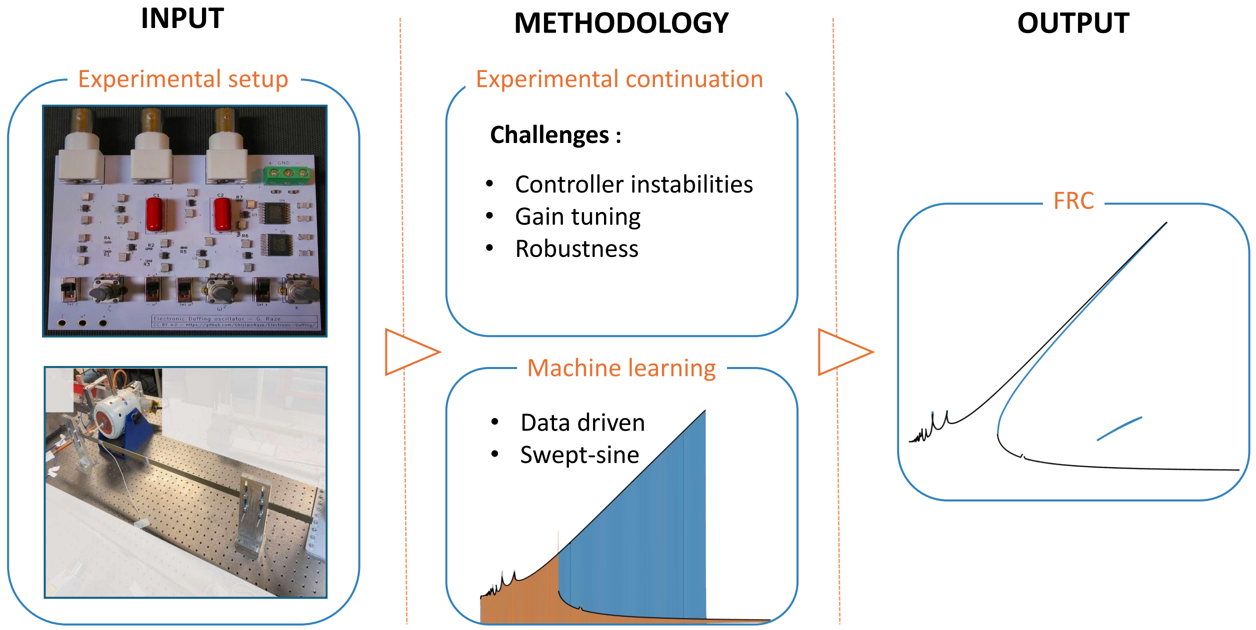


Figure 1.2: Schematic overview of the proposed methodology, from standard input data to machine learning–based prediction of nonlinear frequency response features.

1.4 Outline

This master's thesis is structured into eight chapters.

Chapter 2 introduces the fundamental physical principles of resonance in mechanical systems, contrasting linear and nonlinear dynamics through analytical and numerical exploration of representative oscillators such as the Duffing, Helmholtz-Duffing, and piecewise-linear systems.

Chapter 3 provides a detailed overview of machine learning techniques, with a particular focus on supervised learning and convolutional neural networks. The chapter also discusses the U-Net architecture, which is central to the proposed image-based prediction framework.

Chapter 4 describes the generation of synthetic datasets used for training. It explains the construction of input-output pairs, dataset parameterisation, and the creation of binary image representations that encapsulate essential features of the swept-sine excitation responses.

Chapter 5 applies the trained model to numerical simulations of the Duffing oscillator. The chapter evaluates its performance in predicting both fundamental and secondary (superharmonic and subharmonic) response branches and presents a detailed study of model accuracy under various configurations.

Chapter 6 presents the experimental validation of the methodology using an electronic realisation of the Duffing oscillator. Predictions are compared against ground-truth data obtained from control-based continuation methods, demonstrating the model's ability to infer unstable response branches from swept-sine measurements.

Chapter 7 assesses the generalisation capabilities of the approach by applying it to different classes of nonlinear systems beyond the Duffing oscillator, including Helmholtz-Duffing and piecewise-linear stiffness models. The results highlight the robustness and adaptability of the proposed framework.

Chapter 8 concludes the thesis by summarising the key findings, emphasising the novelty of predicting experimentally inaccessible branches using deep learning, and proposing several directions for future research.

CHAPTER 2

Resonances of linear and nonlinear oscillators

The theoretical and experimental foundations of linear systems are well established. In many engineering applications, linear theory provides accurate predictions as long as deviations between theoretical and experimental results remain small. A key characteristic of linear systems is their mono-harmonic response to a mono-harmonic excitation. In contrast, nonlinear systems exhibit multi-harmonic responses under the same excitation conditions. Additionally, nonlinear effects lead to phenomena such as amplitude-dependent resonances, multi-stable solutions, and jump discontinuities in frequency response [50, 51].

This chapter explores the fundamental differences between linear and nonlinear dynamics through the well-known single-degree-of-freedom (SDOF) Duffing oscillator. First, the undamped and unforced Duffing oscillator, whose exact solution is known, will be compared to its linear equivalent. Next, the harmonically forced and damped Duffing oscillator will be analysed numerically to illustrate the occurrence of multiple resonances, a primary focus of this research.

2.1 Duffing oscillator

The first explored system is the Duffing oscillator [38], a nonlinear oscillator characterised by a cubic stiffness term. The system's nonlinearity arises from its restoring force, where the sign of the cubic term determines whether the oscillator exhibits hardening (positive cubic term) or softening (negative cubic term) behaviour. The Duffing equation has been extensively studied using both numerical and analytical methods, as it captures a wide range of nonlinear dynamical phenomena. Moreover, it serves as a fundamental model for various complex systems, including plates, shells, rotors, and other structural components.

The equation of motion for a harmonically forced Duffing oscillator with linear viscous damping is given by:

$$m\ddot{y}(t) + c\dot{y}(t) + k_1 y(t) + k_3 y^3(t) = F\sin(\omega t), \tag{2.1}$$

where $y(t)$ represents the system's displacement, m is the mass, k_1 and k_3 are the linear and cubic stiffness coefficients, respectively, and c is the linear viscous damping coefficient. The system is subjected to an external excitation of amplitude F and frequency ω.

When normalized with respect to mass, Equation (2.1) transforms into:

$$\ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) + \alpha_3 y^3(t) = \tilde{F}\sin(\omega t), \tag{2.2}$$

where $\omega_0 = \sqrt{k_1/m}$ is the natural frequency, $\tilde{F} = F/m$ is the mass-normalized excitation amplitude, $\zeta = c/(2\sqrt{k_1 m})$ is the damping ratio, and $\alpha_3 = k_3/m$ is the mass-normalized cubic stiffness.

2.1.1 Undamped and unforced Duffing oscillator

Even in the absence of damping and external forcing, linear and nonlinear oscillators exhibit fundamentally different behaviours. The governing equation for both cases is expressed in Equation (2.3), where the nonlinear cubic term, unique to the Duffing oscillator, is highlighted in blue.

$$\begin{cases} \ddot{y}(t) + \omega_0^2\, y(t) + \alpha_3 y^3(t) = 0, \\ y(0) = y_0, \quad \dot{y}(0) = \dot{y}_0. \end{cases} \tag{2.3}$$

For the linear system ($\alpha_3 = 0$), the analytical solution is a harmonic response given by:

$$y(t) = A\sin\left(\omega_0 t - \phi\right), \tag{2.4}$$

where the amplitude of the free oscillation $A = \sqrt{y_0^2 + \left(\frac{\dot{y}_0}{\omega_0}\right)^2}$ and the phase lag $\tan\phi = -\frac{\omega_0\, y_0}{\dot{y}_0}$ depend only on the initial displacement y_0 and velocity $\dot{y}_0$. In contrast, the nonlinear system does not exhibit purely harmonic motion. Instead, the solution for only an initial displacement y_0 is given in terms of the Jacobian elliptic function cn, as shown in Equation (2.5):

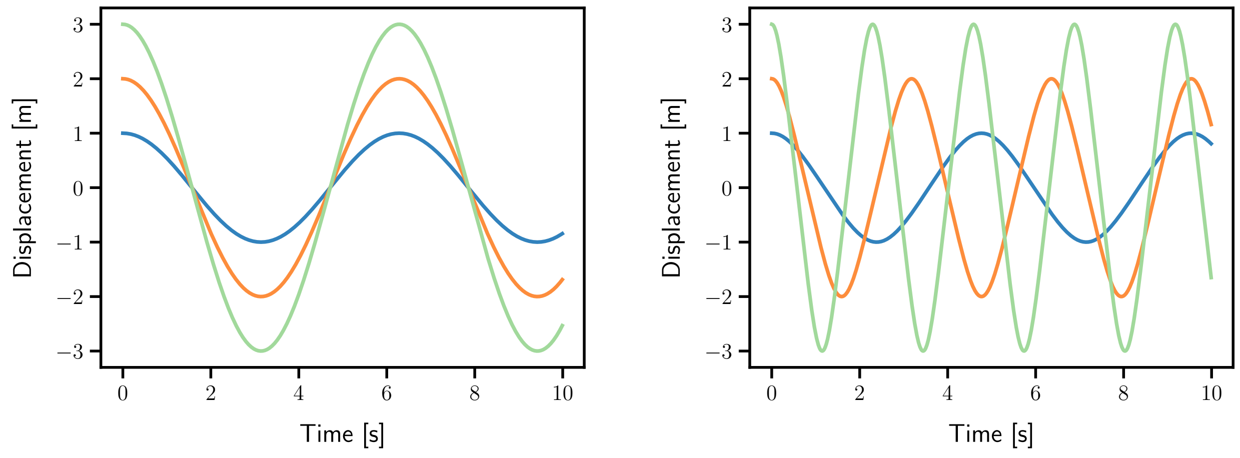
$$y(t) = y_0\, cn\left(\Omega t \,|\, m\right), \tag{2.5}$$

where the nonlinear oscillation frequency Ω and the elliptic parameter m are defined as:

$$\Omega = \sqrt{\omega_0^2 + \alpha_3 y_0^2}, \quad m = \frac{\alpha_3 y_0^2}{2\Omega^2}. \tag{2.6}$$

Unlike the linear case, the frequency Ω now depends on the initial displacement y_0, violating the principle of superposition. This fundamental distinction is illustrated in Figure 2.1, which compares free oscillations of the linear and nonlinear systems under different initial amplitudes y_0.

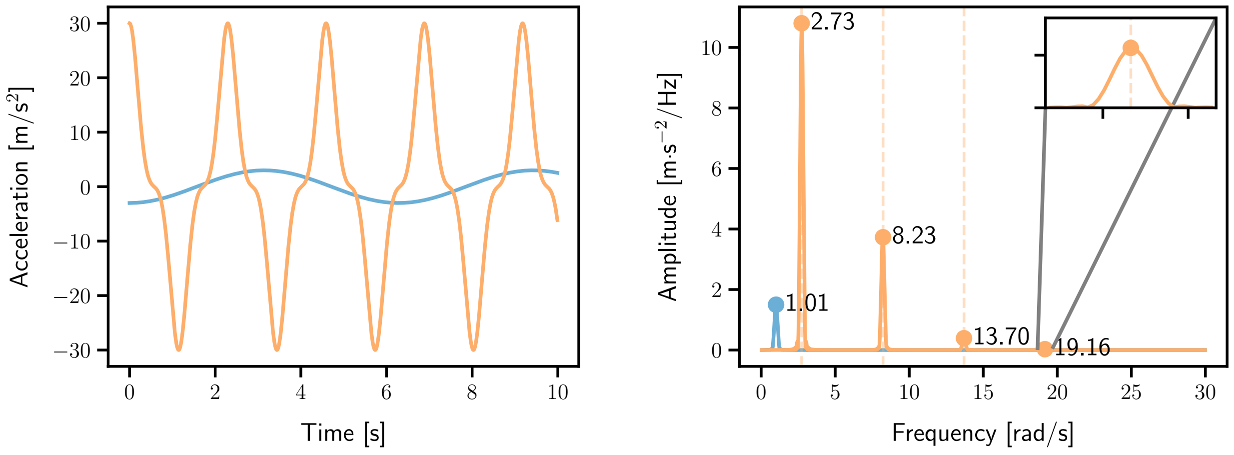
Furthermore, nonlinear oscillations are not purely sinusoidal. The frequency spectrum of the acceleration response, illustrated in Figure 2.2, reveals the presence of higher harmonics beyond the fundamental frequency.



(a) Linear responses.

(b) Nonlinear responses.

Figure 2.1: Comparison of free oscillations for linear and nonlinear systems with initial amplitudes: $y_0 = 1.0$ (—), $y_0 = 2.0$ (—), $y_0 = 3.0$ (—), and $\omega_0 = \alpha_3 = 1.0$.



(a) Time-domain acceleration response.

(b) FFT of the acceleration response.

Figure 2.2: Comparison of accelerations and their frequency spectra for linear (—) and nonlinear (—) systems with initial amplitude $y_0 = 3.0$ and parameters $\omega_0 = \alpha_3 = 1.0$.

2.1.2 Damped and forced Duffing oscillator

In practical applications, systems are subject to external forces and dissipation. To model this, a viscous damping term and a harmonic forcing function are introduced. The governing equation for the linear case is:

$$\begin{cases} \ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) = f\sin(\omega t), \\ y(0) = y_0, \quad \dot{y}(0) = \dot{y}_0. \end{cases} \tag{2.7}$$

Although the system initially depends on y_0 and $\dot{y}_0$, the steady-state solution dictates the long-term response:

$$y_p(t) = \frac{f}{\sqrt{(\omega_0^2 - \omega^2)^2 + (2\zeta\omega_0\omega)^2}} \sin\left(\omega t - \tan^{-1}\left(\frac{2\zeta\omega_0\omega}{\omega_0^2 - \omega^2}\right)\right). \tag{2.8}$$

From Equation (2.8), two key observations can be made:

- The response amplitude peaks near resonance ($\omega \approx \omega_0$).
- The system obeys the principle of superposition, meaning the response scales linearly with the excitation.

The frequency response function of the linear system, commonly visualised using Bode plots, is shown in Figure 2.3.

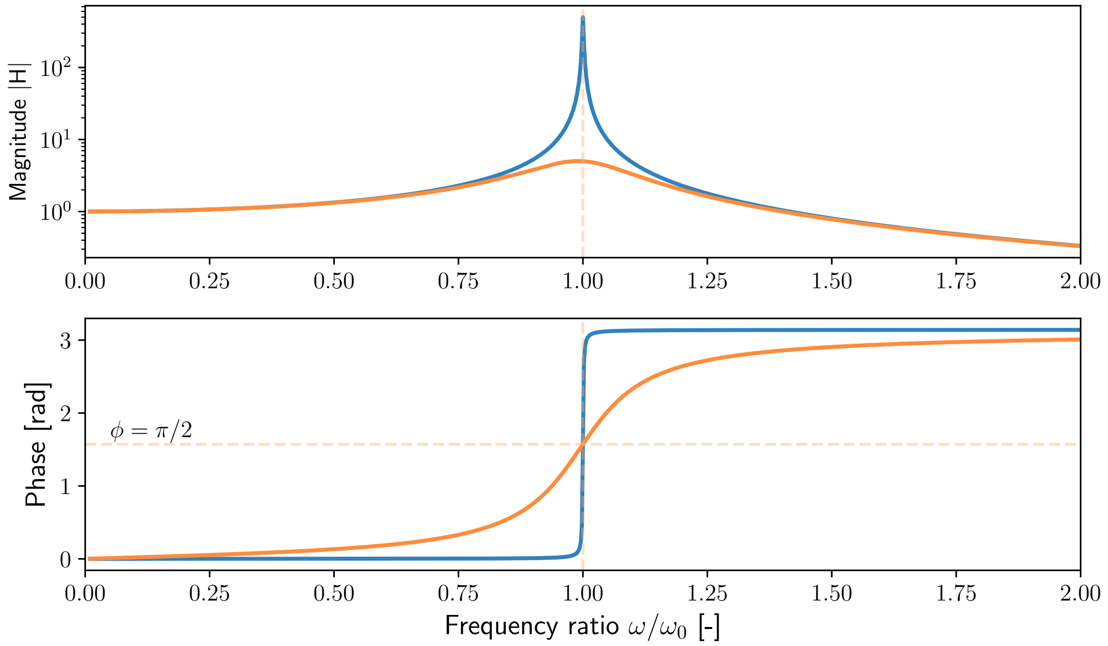


Figure 2.3: Amplitude and phase Bode plots of a damped linear SDOF oscillator for two damping ratios: $\zeta = 0.1\%$ (—) and $\zeta = 10\%$ (—).

The Duffing equation (2.2) can be nondimensionalised to simplify the analysis and highlight the essential parameters governing the system's behaviour. This is accomplished by introducing the dimensionless variables $\tau = \omega_0 t$, $\bar{y} = y/\sigma$, and frequency $\bar{\omega} = \omega/\omega_0$. Substituting these definitions into Equation (2.2) yields:

$$\bar{y}'' + 2\zeta\bar{y}' + \bar{y} + \sigma^2\frac{k_3}{k_1}\bar{y}^3 = \frac{F}{\sigma k_1}\sin(\bar{\omega}\tau). \tag{2.9}$$

Finally, setting $\sigma = \sqrt{k_1/k_3}$ gives the nondimensionalized equation:

$$\bar{y}'' + 2\zeta\bar{y}' + \bar{y} + \bar{y}^3 = \bar{F}\sin(\bar{\omega}\tau), \tag{2.10}$$

where $\bar{F} = F/(\sigma k_1)$ is the nondimensional forcing amplitude.

In contrast to the linear case, where the FRF admits a unique and analytically solvable solution, the forced Duffing oscillator described by Equation (2.10) presents significant challenges due to its nonlinear nature. Due to the inherent nonlinearity of the system, no closed-form solution exists. More importantly, the frequency response may exhibit multiple coexisting solutions, both stable and unstable, for the same excitation frequency. This multi-solution behaviour is a hallmark of nonlinear dynamics and prevents the use of standard linear analysis tools. Consequently, various numerical and analytical techniques are employed to compute the FRCs, including the perturbation method [49, 50], the shooting method [62], and the harmonic balance method [18, 39].

2.1.3 Harmonic balance method

The harmonic balance (HB) method is also known as the Fourier–Galerkin method because it implements the Galerkin method with Fourier basis and test functions. The fundamental concept of this method is that every steady-state response of a nonlinear system, subjected to an excitation frequency ω, can be represented as a finite Fourier series of harmonics of this frequency.

The equations of motion for a nonautonomous nonlinear dynamical system with n degrees of freedom (DOFs) are

$$\mathbf{M}\,\ddot{\mathbf{y}} + \mathbf{C}\,\dot{\mathbf{y}} + \mathbf{K}\,\mathbf{y} + \mathbf{f}_{nl}\left(\mathbf{x}, \dot{\mathbf{x}}\right) = \mathbf{f}_{ext}\left(\omega, t\right) \tag{2.11}$$

where vectors $\mathbf{y}$, $\mathbf{f}_{nl}$, and $\mathbf{f}_{ext}$ represent the displacement, nonlinear forces and external forces, respectively, while $\mathbf{M}$, $\mathbf{C}$ and $\mathbf{K}$ denote the mass, damping and stiffness matrices. The signals $\mathbf{y}(t)$ and $\mathbf{f}\left(\mathbf{y}, \dot{\mathbf{y}}, \omega, t\right) = \mathbf{f}_{nl}\left(\mathbf{y}, \dot{\mathbf{y}}\right) - \mathbf{f}_{ext}\left(\omega, t\right)$ are approximated using a truncated Fourier series up to the N_H-th harmonic:

$$\mathbf{y}(t) = \frac{\mathbf{c}_0^x}{\sqrt{2}} + \sum_{k=1}^{N_H} \left(\mathbf{s}_k^x \sin\left(\omega_k t\right) + \mathbf{c}_k^x \cos\left(\omega_k t\right)\right) \tag{2.12}$$

$$\mathbf{f}(t) = \frac{\mathbf{c}_0^f}{\sqrt{2}} + \sum_{k=1}^{N_H} \left(\mathbf{s}_k^f \sin\left(\omega_k t\right) + \mathbf{c}_k^f \cos\left(\omega_k t\right)\right) \tag{2.13}$$

where $\omega_k = \dfrac{k\omega}{\nu}$ with $\nu \in \mathbb{N}$, $\mathbf{c}_k$ and $\mathbf{s}_k$ are the Fourier series vectors, and the Fourier coefficients $\mathbf{c}_k^x$ and $\mathbf{s}_k^x$ of the displacement become the new unknowns of the problem. The fundamental harmonic corresponds to the term in the series for which $\nu = k$, whereas the secondary harmonics include all components for which $\nu \neq k$. Ultimately, the displacement $\mathbf{y}(t)$, velocities $\dot{\mathbf{y}}(t)$, accelerations $\ddot{\mathbf{y}}(t)$ (all defined using Fourier series), and forces $\mathbf{f}(t)$ are substituted into the equations of motion 2.11. This yields a set of $2N_H + 1$ nonlinear algebraic equations solved iteratively, typically with a Newton-Raphson scheme.

Notably, for a given forcing level, a nonlinear system can exhibit multiple solutions corresponding to phenomena such as low- and high-amplitude vibrations. Consequently, depending on the initial guess provided to the Newton-Raphson procedure, the algorithm will converge toward one of these solutions. To compute additional solutions, methods based on homotopy [45], such as a two-harmonic homotopy method (THHM) [61], will be employed in this research for the computation of subharmonic responses.

One of the key strengths of the HB method lies in its accuracy, even with low-order approximations. While incorporating more harmonics enhances solution precision, it inevitably increases computational resources. In applications involving a large number of computations, the HB method parameters are selected to optimally balance accuracy and computational cost.

The FRC of the harmonically forced and damped nonlinear SDOF nondimensionalised oscillator (see Equation 2.10) is computed using HB with pseudo-arclength continuation [5, 19] and the FRCs for two force levels are depicted in Figure 2.4. This example illustrates well the differences with linear systems, it is no longer a function, and there is a clear nonlinear dependency between the response and forcing level.

The presence of multi-stable solutions, typically associated with fold (saddle-node) bifurcations, arises when the derivative of the amplitude with respect to the excitation frequency vanishes. This phenomenon can induce a jump effect [50], where the system abruptly transitions from one stable solution branch to another as a control parameter, such as the excitation frequency is varied. The intermediate branch connecting these bifurcation points is dynamically unstable. Consequently, this unstable branch cannot be captured using classical linear or frequency sweep methods. Instead, specialised approaches such as control-based continuation or phase-locked loop techniques are required to systematically track these solutions, as discussed in the introductory section.

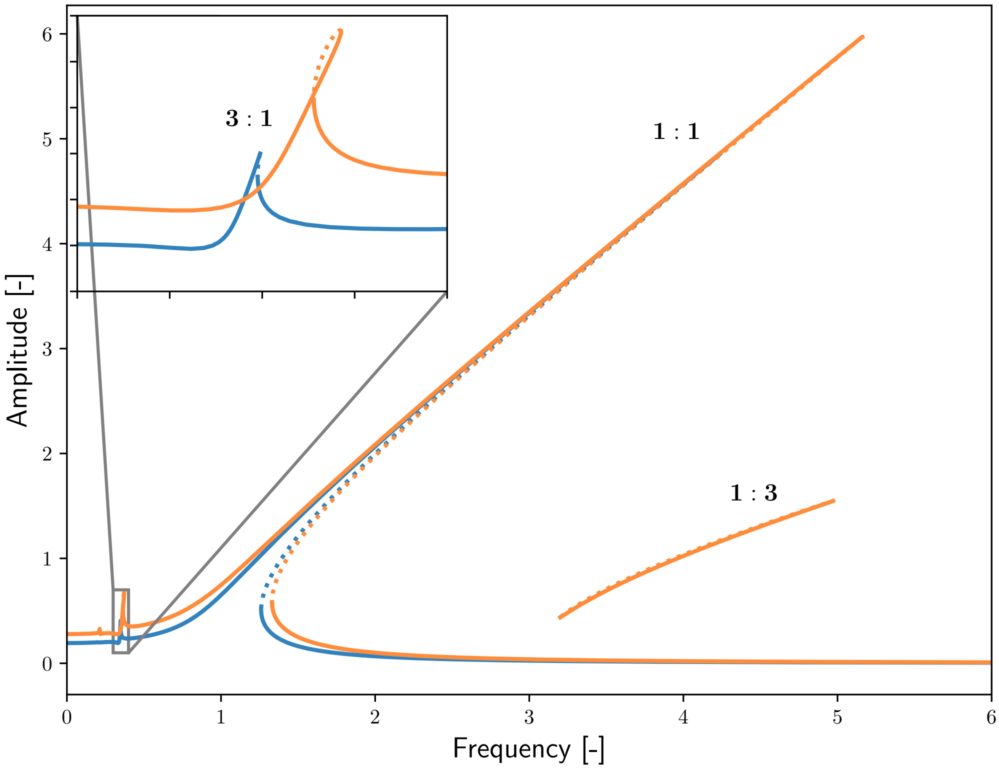


Figure 2.4: Nonlinear frequency response curves of the Duffing oscillator for two forcing levels: $\bar{F} = 0.2$ (—) and $\bar{F} = 0.3$ (—) with a damping ratio $\zeta = 0.5\%$. Stable and unstable branches are represented by solid and dotted lines, respectively.

In nonlinear systems, each harmonic l of the response's Fourier series can trigger a resonance if $\omega_k = k\,\omega/\nu$ corresponds to the frequency of the system's primary resonance. According to Stoker [73], resonances can be classified into four distinct categories:

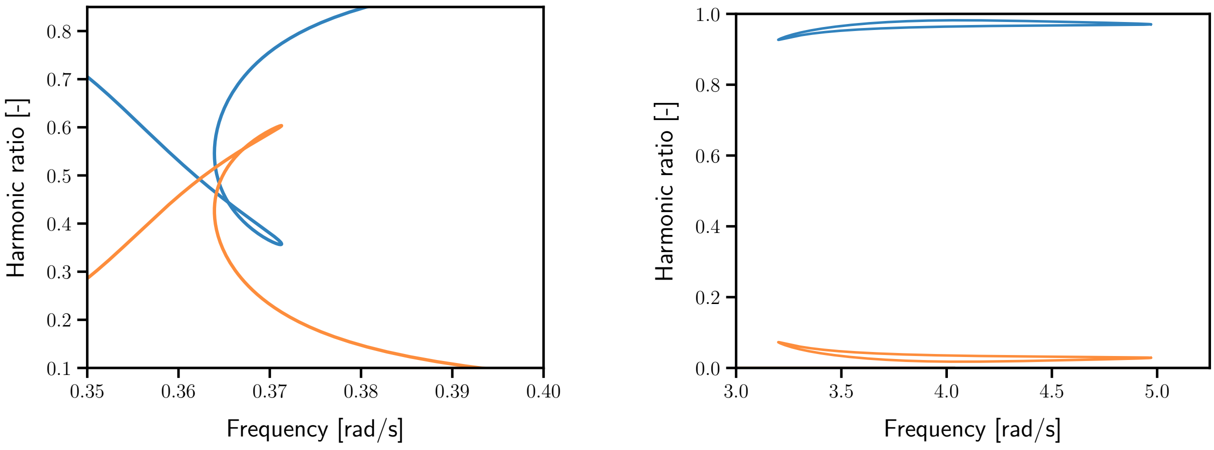
- 1:1 primary/fundamental resonance ($k = \nu = 1$)
- k:1 superharmonic or ultraharmonic resonances
- 1:ν subharmonic resonances

- k:ν ultra-subharmonic resonances

This work focuses on predicting the unstable branches of the fundamental, superharmonic, and complete subharmonic resonances while excluding the prediction of ultra-subharmonic resonances.

In the example shown in Figure 2.4, a small peak is observed around $\omega_0/3$ for both forcing levels. This peak corresponds to the 3:1 superharmonic resonance. Figure 2.5a shows the harmonic ratios for the first and third harmonics around the 3:1 superharmonic resonance. As expected, the harmonic ratio of the third harmonic increases around $3\,\omega \approx \omega_0$, highlighting the superharmonic resonance mechanism.

However, when the forcing amplitude is increased to $\bar{F} = 0.3$, an additional isolated response branch appears beyond $3\,\omega_0$, detached from the main resonance. This corresponds to the 1:3 subharmonic resonance. The inherent complexity of nonlinear systems is evinced by the observation that even a modest increase in forcing amplitude can induce a qualitatively distinct system response. The first harmonic component dominates in this frequency range, with the third harmonic being almost absent compared to the first harmonic, as evidenced by the significant difference in magnitude between these harmonics in the subharmonic resonance region as shown in Figure 2.5b.



(a) 3:1 superharmonic

(b) 1:3 subharmonic

Figure 2.5: Harmonic ratios ($\bar{F} = 0.3$ and $\zeta = 0.5\%$) around the resonances ((—): $l = 1$; (—): $l = 3$).

As the forcing amplitude increases further to $\bar{F} = 1.0$, a rich spectrum of additional secondary resonances emerges, as illustrated in Figure 2.6. These resonances appear both above and below the fundamental resonance frequency. The new resonances include both odd and even superharmonics where $\nu = 1$ and l vary (2:1, 3:1, 4:1, 5:1, 7:1, etc.). Additionally, another isolated branch appears in the frequency response, corresponding to the 2:1 subharmonic resonance.

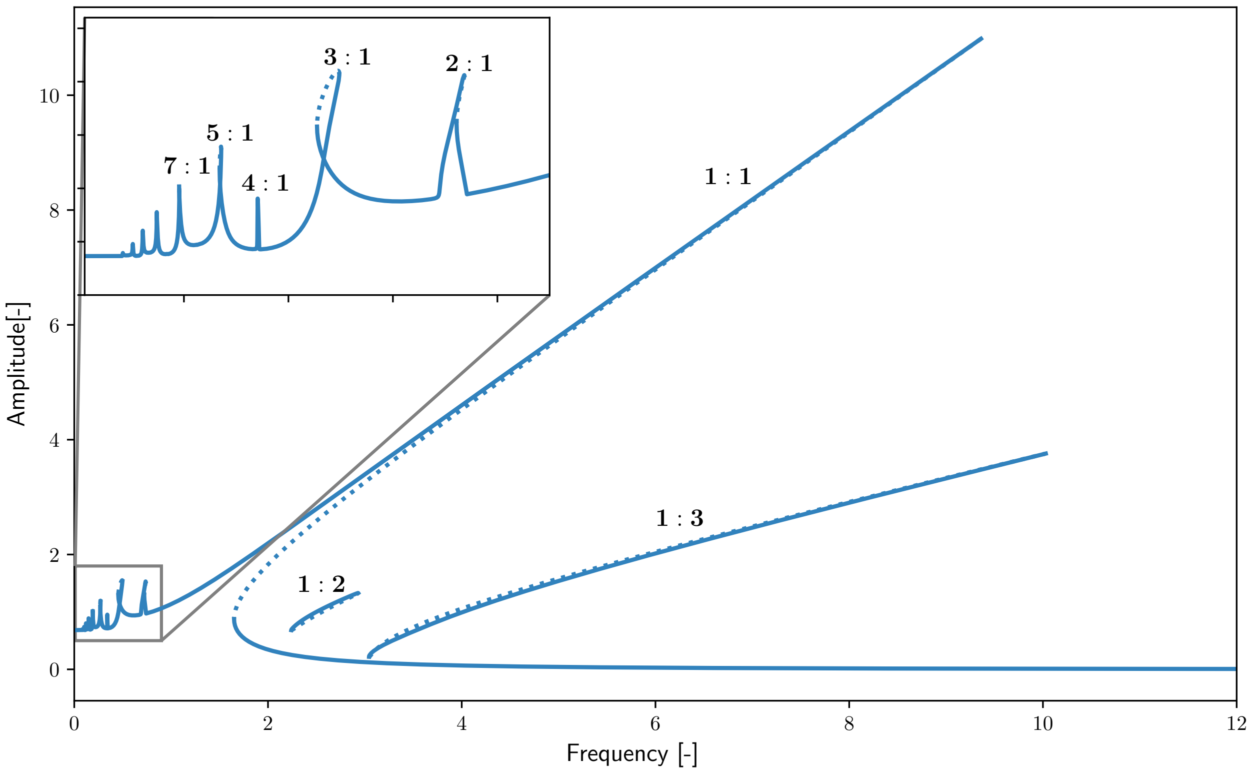


Figure 2.6: FRC of the Duffing oscillator with $\bar{F} = 1.0$ and the damping ratio $\zeta = 0.5\%$. Stable and unstable branches are represented by solid and dotted lines, respectively.

2.2 Swept-sine excitation

Several well-established methods exist for experimentally obtaining the stable branches of an FRC. The swept-sine (or sine-sweep) testing is one of the most widely used methods, valued for both its simplicity and effectiveness. The swept-sine test can begin at a low frequency and then sweep to a high frequency (swept up) or vice versa (swept down).

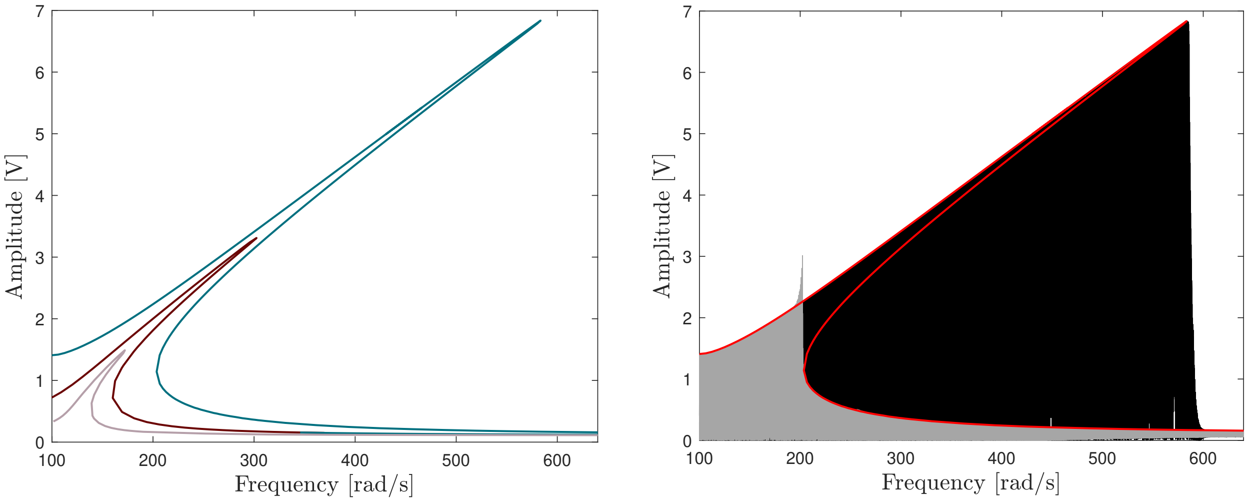
The swept-sine excitation corresponds to a sinusoidal excitation applied to the system while the excitation frequency is gradually varied. The force equation of a linear swept-sine is:

$$f_{\text{ext}}(\omega, t) = \text{A} \ \sin\left(\pi\left[(f_2 - f_1)\left(\frac{t}{T}\right) + f_1\right]\right), \tag{2.14}$$

where A is the excitation amplitude, f_1 and f_2 are, respectively, the starting and ending frequencies, and T is the test duration. This signal is fully deterministic and very easy to interpret visually. Furthermore, it strongly activates the nonlinearities, making it an excellent choice for nonlinearity characterisation [21]. When greater resolution at lower frequencies is desired, a logarithmic sine sweep can be implemented.

A significant advantage of swept-sine testing is its practical implementation, it is uncomplicated to perform, fast to execute, and requires only the adjustment of a single parameter: the sweep rate. This simplicity makes it accessible for numerous applications in experimental mechanics.

With swept-sine testing, the jump phenomenon occurs between coexisting stable solutions, exposing a fundamental limitation of open-loop experimental modal analysis (EMA), which is not designed to characterise nonlinear dynamics where identical excitation conditions can yield multiple responses. What is typically observed under such excitation is a sudden transition, or jump, from one stable branch to another, occurring near a fold (saddle-node) bifurcation. This jump highlights the presence of the bifurcation, which forms the boundary between stable and unstable solution branches. By performing both forward and backward frequency sweeps, one can capture hysteresis effects and approximate the location of these folds, as illustrated in Figure 2.7.



(a) **(b)**

Figure 2.7: (a) FRCs around the fundamental resonance of the electronic Duffing oscillator for distinct forcing levels. (b) Comparison of the FRC (red) at $\bar{F} = 2\,\mathrm{V}$ with swept-up (black) and swept-down (grey) sine tests. [72]

However, in practical tests, the observed jump may deviate from the actual fold (saddle-node) bifurcation point occurring earlier due to external noise or later as a result of finite sweep rates and associated dynamic bifurcation effects [36].

2.3 Helmholtz-Duffing and piecewise linear stiffness oscillators

This section introduces two other important nonlinear oscillator types: the Helmholtz-Duffing oscillator, characterised by its cubic and quadratic stiffness terms, and the piecewise linear stiffness oscillator.

2.3.1 Helmholtz-Duffing oscillator

The Helmholtz-Duffing oscillator extends the standard Duffing model by incorporating an additional quadratic nonlinearity. Its mass-normalised equation of motion is expressed as:

$$\ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) + \alpha_2 y^2(t) + \alpha_3 y^3(t) = \gamma \sin(\omega t), \tag{2.15}$$

where all parameters retain their previous definitions, except for $\alpha_2 = k_2/m$, which denotes the mass-normalized quadratic stiffness coefficient.

Applying the same non-dimensionalisation procedure used for the Duffing oscillator yields:

$$\bar{y}'' + 2\zeta\bar{y}' + \bar{y} + \beta\bar{y}^2 + \bar{y}^3 = \bar{F}\sin(\bar{\omega}\tau) \tag{2.16}$$

where $\beta = \dfrac{\alpha_2\sqrt{\alpha_3}}{\omega_0}$ is the non-dimensional quadratic stiffness parameter. The same parameters thus characterise the system as the Duffing oscillator with the addition of $\beta\bar{y}^2$.

Figure 2.8 illustrates the FRC and phase resonance nonlinear mode (PRNM) curves for this oscillator with parameters $\zeta = 0.5\%$, $\beta = 1.2$, and $\bar{F} = 0.2$.

The Helmholtz-Duffing oscillator is often referred to as a softening-hardening oscillator due to its dual nonlinear behaviour. At low forcing amplitudes, the quadratic nonlinearity ($\beta\,\bar{y}^2$ term) produces a softening effect, causing the resonance peak to bend toward lower frequencies. At higher amplitudes, the cubic nonlinearity ($\bar{y}^3$ term) dominates, resulting in a hardening effect that shifts the resonance peak toward higher frequencies. This transition between softening and hardening behaviours is particularly evident in the PRNM curves of the different resonances shown in Figure 2.8.

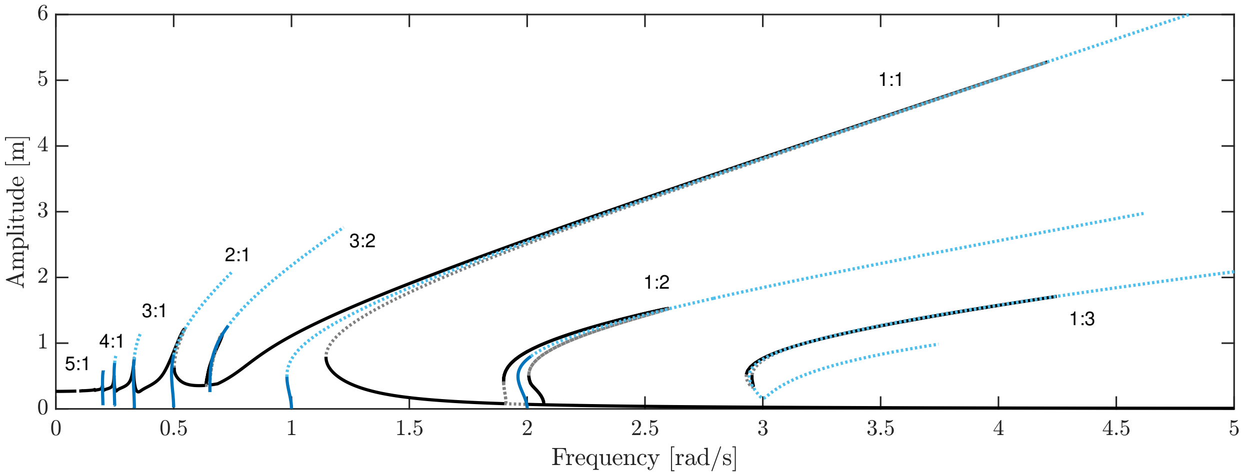


Figure 2.8: Nonlinear frequency response curves (black: stable; grey: unstable) and phase resonance nonlinear mode (PRNM) curves (solid blue: stable; dotted blue: unstable) of the Helmholtz-Duffing oscillator. [78]

Piecewise linear stiffness oscillator

Piecewise linear stiffness (PWL) oscillators represent an important class of nonlinear systems with wide-ranging applications in structural dynamics, vibration isolation, and impact mechanics [51, 67]. This section examines the SDOF oscillator with viscous damping and a piecewise linear stiffness nonlinearity, described by:

$$m\ddot{y}(t) + c\dot{y}(t) + k_1 y(t) + f_{\mathrm{nl}}(y(t)) = F\sin(\omega t), \tag{2.17}$$

Where $f_{\mathrm{nl}}(y(t))$ represents the nonlinear restoring force characterised by a piecewise linear function:

$$f_{\mathrm{nl}}(y) = \begin{cases} k_{\mathrm{pwl}}(y + y_{\mathrm{lim}}), & y \le -y_{\mathrm{lim}}, \\ 0, & -y_{\mathrm{lim}} < y < y_{\mathrm{lim}}, \\ k_{\mathrm{pwl}}(y - y_{\mathrm{lim}}), & y \ge y_{\mathrm{lim}}. \end{cases} \tag{2.18}$$

The parameter y_{lim} represents the displacement threshold beyond which the additional stiffness k_{pwl} becomes active. This configuration is frequently encountered in mechanical systems with clearances, impact limiters, or elastic constraints [31, 53].

The nondimensionalization of the system is performed by introducing the normalised displacement $\bar{y} = y/y_{\text{lim}}$ and the dimensionless time variable $\tau = \omega_0 t$. Substituting these expressions into the equation of motion and dividing through by $k_1 y_{\text{lim}}$ yields the non-dimensional form of the governing equation:

$$\bar{y}'' + 2\zeta\bar{y}' + \bar{y} + \bar{f}_{\text{nl}}(\bar{y}) = \bar{F}\sin(\bar{\omega}\tau), \tag{2.19}$$

where primes denote differentiation with respect to τ, and the nondimensional restoring force is expressed as:

$$\bar{f}_{\text{nl}}(\bar{y}) = \begin{cases} \kappa(\bar{y}+1), & \bar{y} \leq -1, \\ 0, & -1 < \bar{y} < 1, \\ \kappa(\bar{y}-1), & \bar{y} \geq 1. \end{cases} \tag{2.20}$$

In practical applications, this piecewise-linear model is often regularised by introducing a smooth polynomial transition (typically quadratic or cubic) in the vicinity of $\bar{y} = \pm 1$ to ensure continuity and differentiability of the restoring force. The dimensionless parameters governing the system response are defined as:

$$\kappa = \frac{k_{\text{pwl}}}{k_1}, \qquad \bar{F} = \frac{F}{k_1 y_{\text{lim}}}. \tag{2.21}$$

Here, κ is the stiffness ratio between the additional piecewise linear stiffness k_{pwl} and the linear stiffness k. This formulation yields a system characterised by three independent parameters: ζ, $\bar{F}$, and κ.

Figure 2.9 presents the FRCs of the piecewise linear oscillator at $\bar{F} = 1$ and $\zeta = 5\%$ for $\kappa = 5$ and $\kappa = 10$. Compared to the Duffing and Helmholtz-Duffing cases [38], the non-smooth nonlinearity produces markedly more intricate resonance features [10].

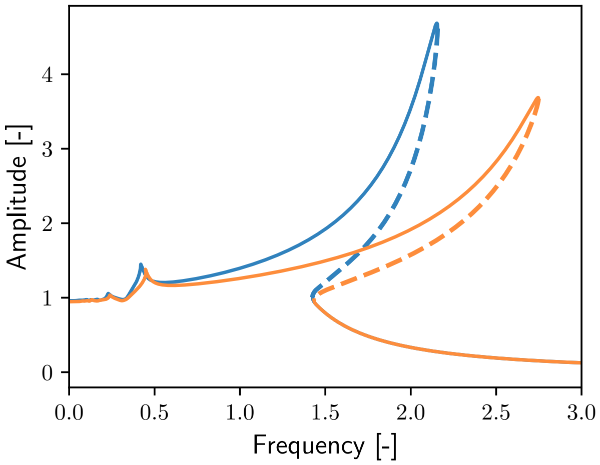


Figure 2.9: Frequency response curves of the piecewise linear oscillator for $\bar{F} = 1$ and $\zeta = 5\%$. (—): $\kappa = 5$; (—): $\kappa = 10$. Solid and dashed lines are, respectively, stable and unstable branches of solution.

2.4 Conclusion and discussion on resonances

Chapter conclusion

This chapter highlights the fundamental differences between resonances in linear and nonlinear systems, examining them through the Duffing oscillator using both analytical and numerical approaches.

Starting with the undamped and unforced linear and nonlinear oscillators, the analytical solutions revealed several key distinctions between the two systems. Nonlinear systems exhibit non-pure harmonic motion, the presence of multiple harmonics, and a violation of the principle of superposition that governs linear systems.

The resonance characteristics of nonlinear systems demonstrated significantly greater complexity, as illustrated in Figures 2.4 and 2.6. The primary resonance shifts to considerably higher frequencies due to a hardening effect. Additionally, the presence of multi-stable solutions can lead to jump phenomena, potentially causing dramatic amplitude changes over small frequency variations, which may result in system damage [48, 81].

Furthermore, multiple peaks observed before and after the fundamental resonance stem from the multi-harmonic nature of the system response, creating secondary resonances. The $l : 1$ superharmonics with odd l were computed using classical harmonic balance continuation, while the $l : 1$ superharmonics with even l bifurcate from the main branch, necessitating more sophisticated bifurcation analysis techniques [18, 20]. The remaining resonances appear as isolated branches of solutions, which can be revealed through basins of attraction calculations or, as employed in this research, through the two-harmonic homotopy method [61].

In addition to the analysis of the Duffing oscillator, swept-sine excitation was employed to illustrate the system's resonant behaviour experimentally. This technique not only confirmed the numerical predictions but also highlighted hysteresis effects and the jump phenomenon arising from fold bifurcations.

The study was further extended to include the Helmholtz-Duffing oscillator, which incorporates both quadratic and cubic nonlinearities. This dual nonlinearity produces a characteristic transition from softening to hardening behaviour as the excitation amplitude increases, thereby shifting the resonance peak and introducing additional complexity into the system's response.

Moreover, a piecewise linear stiffness oscillator was investigated to capture the effects of non-smooth nonlinearities.

Overall, this chapter offers valuable insights into the complex dynamics of nonlinear resonant systems, presenting a comprehensive approach that combines analytical and numerical methodologies.

CHAPTER 3

Machine learning for nonlinear frequency response prediction

Machine learning (ML), a subfield of artificial intelligence (AI), develops systems that learn to make decisions by fitting mathematical models to observed data. A deep neural network represents a specific class of machine learning models, where the process of fitting these models to data is referred to as deep learning (DL).

This chapter provides an introduction to deep learning with a specific emphasis on its application to the present research. The conceptual framework primarily draws upon the principles discussed in *Understanding Deep Learning* by Simon J.D. Prince [58].

Machine learning methodologies are conventionally categorised into three distinct paradigms: supervised learning, unsupervised learning, and reinforcement learning.

Supervised learning aims to establish a mapping function from input data to output predictions. These input/output pairs can manifest in various forms and dimensions, for which specialised DL architectures have been developed. While the fundamental objective remains consistent, mapping inputs to outputs, the nature of the output varies by problem type. In regression problems, the model generates continuous numerical values, whereas, in classification problems, a binary classifier assigns inputs to one of two categories or to one of N categories (where $N > 2$) in multiclass classification scenarios.

Unsupervised learning adopts a fundamentally different approach. Rather than learning input-output mappings, it seeks to discern inherent structures within unlabeled data. Models are trained solely on input data without corresponding output labels. Principal unsupervised learning techniques include clustering, dimensionality reduction, and association rule mining [12]. Contemporary state-of-the-art unsupervised models encompass generative architectures [76], which can synthesise highly plausible yet novel samples that are distinct from the training data.

Reinforcement learning operates on a different paradigm centred around an agent executing actions at sequential time steps. Each action potentially yields a reward, with the agent's objective being to learn strategies that maximise cumulative rewards. This framework introduces the exploration-exploitation dilemma [75], where the agent must balance discovering new potentially rewarding actions against exploiting known rewarding actions.

This research introduces a supervised learning framework to predict the unstable branches of FRCs, a novel application of deep learning in this context. While this method offers a new perspective, it is essential to note that various other approaches exist for analysing nonlinear dynamics [28, 80, 82], although none have specifically addressed the prediction of FRC as undertaken here.

3.1 Supervised learning

Supervised learning aims to construct a model $\mathbf{f}[\cdot]$ parameterized by $\boldsymbol{\phi}$, which maps an input $\mathbf{x}$ to an output $\mathbf{y}$, such that:

$$\mathbf{y} = \mathbf{f}[\mathbf{x}, \boldsymbol{\phi}] \tag{3.1}$$

The process of computing the prediction $\mathbf{y}$ from an input $\mathbf{x}$ using the trained model is called inference. While inference is computationally straightforward, the challenge lies in finding the optimal model parameters.

The learning or training phase of a model involves determining the optimal parameters that map each training input to its associated output with maximum accuracy. This optimization process utilizes a training dataset comprising I input-output pairs $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{I}$. The discrepancy between the model's predictions and the true outputs is quantified by a loss function $L[\boldsymbol{\phi}]$, which measures how accurately the model predicts the training outputs from their corresponding inputs for a given parameter set $\boldsymbol{\phi}$.

The training phase can thus be formulated as an optimisation problem seeking parameters $\hat{\boldsymbol{\phi}}$ that minimise the loss function:

$$\hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi}}{\operatorname{argmin}} \; L\left[\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^{I}, \boldsymbol{\phi}\right]. \tag{3.2}$$

After minimising the loss function, the model parameters are computed such that the training outputs $\mathbf{y}_i$ can be accurately predicted from the training inputs $\mathbf{x}_i$. However, the ultimate objective of supervised learning extends beyond memorising the training data. The true measure of a model's efficacy lies in its ability to generalise to previously unseen examples. Therefore, after training, the model's performance must be rigorously evaluated on a separate test dataset comprising samples not used during training. This evaluation quantifies the model's generalisation capability, a fundamental criterion for its practical utility in real-world applications.

3.2 Loss function

A loss function or cost function $L[\boldsymbol{\phi}]$ returns a single value describing the mismatch between the model predictions $\mathbf{f}[\mathbf{x}_\mathrm{i}, \boldsymbol{\phi}]$ and the corresponding ground-truth outputs $\mathbf{y}_i$. The choice of the loss function is highly problem-dependent and plays a critical role in the performance and training dynamics of machine learning models. An inappropriate loss function can lead to suboptimal convergence, poor generalisation, or even complete training failure. As highlighted by Dräger and Dunkelau [22], the loss function acts as a central hyperparameter whose selection significantly influences the quality of the learned model.

A model actually computes the conditional probability distribution $\Pr(\mathbf{y}|\mathbf{x})$ over possible outputs $\mathbf{y}$ given input $\mathbf{x}$. The loss encourages each training output $\mathbf{y}_i$ to have a high probability under the distribution $\Pr(\mathbf{y}|\mathbf{x})$ computed from the corresponding input $\mathbf{x}_i$. The maximum likelihood criterion aims to find the set of parameters $\boldsymbol{\phi}$ that maximises the joint probability of observing all I training examples. This can be expressed as:

$$\hat{\boldsymbol{\phi}} = \underset{\boldsymbol{\phi}}{\text{argmax}} \left[\prod_{i=1}^{I} Pr(\mathbf{y}_i|\mathbf{f}[\mathbf{x}_i, \boldsymbol{\phi}]) \right]. \tag{3.3}$$

As the product of conditional probability distribution, which can be small, can lead to difficulties in representing it with finite precision arithmetic, the logarithm of the likelihood is used. Furthermore, model fitting problems are, by convention, formulated as a minimisation of a loss, which gives the negative log-likelihood criterion

$$\hat{\boldsymbol{\phi}} = \underset{\phi}{\text{argmin}} \left[-\sum_{i=1}^{I} \log\left[Pr(\mathbf{y}_i|\mathbf{f}[\mathbf{x}_i, \boldsymbol{\phi}]) \right] \right] = \underset{\phi}{\text{argmin}} \; [L[\boldsymbol{\phi}]], \tag{3.4}$$

forming the final loss function $L[\boldsymbol{\phi}]$. Furthermore, when performing inference, a point estimate is returned instead of the distribution, i.e. the maximum of the distribution :

$$\hat{\mathbf{y}} = \underset{\text{y}}{\text{argmax}} \left[Pr(\mathbf{y}|\mathbf{f}[\mathbf{x}, \hat{\boldsymbol{\phi}}]) \right]. \tag{3.5}$$

3.2.1 Binary classification

Binary classification aims to assign input data $\mathbf{x}$ to one of two discrete classes $y \in \{0, 1\}$. For this task, the Bernoulli distribution provides an appropriate probabilistic framework, as it is defined precisely on the binary domain $y \in \{0, 1\}$. The Bernoulli distribution is characterized by a single parameter $\lambda \in [0, 1]$ representing the probability that y takes the value 1:

$$Pr(y|\lambda) = \begin{cases} 1 - \lambda, & \text{if } y = 0 \\ \lambda, & \text{if } y = 1 \end{cases} = (1 - \lambda)^{1-y} \lambda^{y}. \tag{3.6}$$

To generate valid probability values, the neural network's output must be transformed from the real domain $\mathbb{R}$ to the interval $[0, 1]$, aligning with the definition of λ. The logistic sigmoid function (see Figure 3.1) serves as an ideal transformation for this purpose:

$$\text{sig}[z] = \frac{1}{1 + \exp[-z]} \tag{3.7}$$

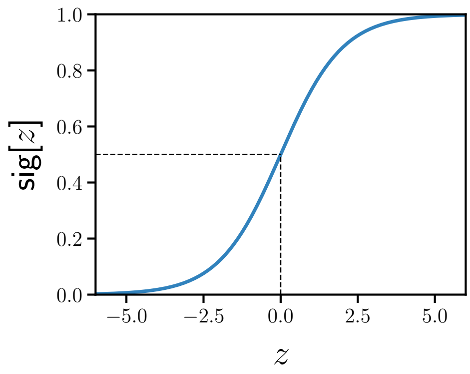


Figure 3.1: The logistic sigmoid function, used to map real-valued inputs to the interval [0, 1].

The sigmoid function's output can be directly interpreted as the probability parameter λ of the Bernoulli distribution, enabling the model to express its confidence in the classification decision. When performing inference, a point estimate may be desired, such that it is set that y $= 1$ if $\lambda >$ threshold τ, and y = 0; otherwise, the threshold is often set to 0.5. The likelihood for an individual sample can be expressed as:

$$Pr(y|\mathbf{x}, \boldsymbol{\phi}) \;=\; (1 - \mathrm{sig}[\mathrm{f}[\mathbf{x}, \boldsymbol{\phi}]])^{1-y} \cdot \mathrm{sig}[\mathrm{f}[\mathbf{x}, \boldsymbol{\phi}]]^{y} \tag{3.8}$$

The loss function is then defined as the negative log-likelihood across the entire training set:

$$L[\boldsymbol{\phi}] \;=\; -\sum_{i=1}^{I} \Big\{ (1 - y_i) \log\left[1 - \mathrm{sig}[\mathrm{f}[\mathbf{x}_i, \boldsymbol{\phi}]]\right] \;+\; y_i \log\left[\mathrm{sig}\left[\mathrm{f}\left[\mathbf{x}_i, \boldsymbol{\phi}\right]\right]\right] \Big\}, \tag{3.9}$$

known as the binary cross-entropy (BCE) loss. The BCE loss is the standard approach for binary classification tasks; however, it becomes suboptimal in scenarios with significant class imbalance. When one class substantially outnumbers the other, standard maximum likelihood estimation becomes inefficient. Under such conditions, the model tends to develop overconfidence in predicting the dominant class while underperforming the minority class.

To address this limitation, Lin et al. introduced the focal loss (FL) [46], specifically designed to handle class imbalance in classification problems:

$$FL[\boldsymbol{\phi}] \;=\; -\sum_{i=1}^{I} \Big\{ \alpha_t (1 - p_t)^{\gamma} \log(p_t) \Big\}, \tag{3.10}$$

where p_t is defined as:

$$p_t = \begin{cases} \mathrm{sig}[\mathrm{f}[\mathbf{x}_i, \boldsymbol{\phi}]], & \text{if } y_i = 1 \\ 1 - \mathrm{sig}[\mathrm{f}[\mathbf{x}_i, \boldsymbol{\phi}]], & \text{if } y_i = 0 \end{cases} \tag{3.11}$$

The focal loss introduces two crucial parameters: α and γ, which must be carefully tuned as hyperparameters:

- α_t is a class-balancing weight that compensates for the frequency disparity between classes. Typically, α is set inversely proportional to class frequency, with $\alpha_t = \alpha$ for $y_i = 1$ and $\alpha_t = 1-\alpha$ for $y_i = 0$.

- $\gamma \geq 0$ is the focusing parameter that modulates the rate at which easy examples are down-weighted. Higher values of γ increase the emphasis on hard, misclassified examples. When $\gamma = 0$, the focal loss reduces to the weighted cross-entropy.

These parameters must be determined through cross-validation or hyperparameter optimisation. Typical values range from 0.25 to 0.75 for α and 1 to 5 for γ, but optimal settings depend on the dataset's class distribution and complexity.

3.3 Fitting models and gradient descent

The network parameters $\boldsymbol{\phi}$ minimising the loss are obtained through a gradient descent iterative procedure:

Step 1: Compute loss gradient: $\frac{\partial L}{\partial \boldsymbol{\phi}} = \left[\frac{\partial L}{\partial \phi_0} \;\; \cdots \;\; \frac{\partial L}{\partial \phi_N} \right]^T$

Step 2: Update parameters: $\boldsymbol{\phi} \leftarrow \boldsymbol{\phi} - \alpha \cdot \frac{\partial L}{\partial \boldsymbol{\phi}}$

Loss functions for nonlinear models are typically non-convex, containing multiple local minima and a global minimum. Random initialisation of $\boldsymbol{\phi}$ provides no guarantee that gradient descent will converge to the global minimum. Additionally, saddle points in the loss may prematurely terminate the optimisation process.

Neural networks with millions of parameters render standard gradient descent suboptimal for finding the global minimum in high-dimensional spaces. Stochastic gradient descent (SGD) addresses this limitation by introducing noise at each iteration. Rather than computing gradients across the entire dataset simultaneously, SGD randomly samples subsets (minibatches) without replacement.

The general update rule can be expressed as:

$$\boldsymbol{\phi}_{t+1} \leftarrow \boldsymbol{\phi}_t - \alpha \cdot \nabla_{\boldsymbol{\phi}} L_t \tag{3.12}$$

where for standard gradient descent $\nabla_{\boldsymbol{\phi}} L_t = \dfrac{\partial L}{\partial \boldsymbol{\phi}}$ computed over all samples, and for SGD $\nabla_{\boldsymbol{\phi}} L_t = \sum_{i \in \mathcal{B}_t} \dfrac{\partial l_i[\boldsymbol{\phi}_t]}{\partial \boldsymbol{\phi}}$ computed over minibatch $\mathcal{B}_t$.

This stochasticity enables temporary uphill movements, facilitating escape from local minima by allowing transitions between different loss function valleys. An epoch constitutes one complete pass through the training dataset.

SGD is frequently implemented with learning rate scheduling, which dynamically adjusts the learning rate α throughout training. Initially, larger values facilitate broad exploration of the parameter space, while subsequent reductions enable fine-tuning through smaller parameter adjustments.

A significant enhancement to SGD incorporates a momentum term in the update rule. Parameters $\boldsymbol{\phi}$ are updated using a weighted combination of the current batch gradient and the previous update direction. Adaptive Moment Estimation (Adam) extends this concept by applying momentum to the estimate of the gradient and the squared gradient. This optimisation algorithm introduces two additional hyperparameters, β_1 and β_2, which control the exponential decay rates for the moment estimates. Adam demonstrates reduced sensitivity to learning rate initialisation [37].

The update of parameters $\boldsymbol{\phi}$ necessitates the computation of loss derivatives. Contemporary state-of-the-art models contain up to $\sim 10^{12}$ parameters, underscoring the critical requirement for efficient derivative calculation. The backpropagation algorithm provides an optimal computational framework for determining these derivatives through the systematic application of the chain rule [65], enabling feasible training of deep neural architectures despite their massive parametric complexity.

3.4 Convolutional networks

When processing images, fully connected networks generate an excessive number of parameters due to their dense connectivity. Furthermore, they fail to exploit the statistical correlations between adjacent pixels and lack mechanisms to maintain consistent feature detection across spatial locations.

Convolutional layers address these limitations by processing local receptive fields with shared filter parameters across the entire spatial domain. This architecture significantly reduces parameter count while preserving sensitivity to local spatial structures. Crucially, convolutional layers exhibit translational equivariance, a property ensuring that spatial shifts in the input produce corresponding shifts in the output features while preserving their values. This means features (e.g., edges, textures, or objects) are detected consistently regardless of their position in the image, eliminating the need to learn separate representations for the same feature at different locations. A network primarily composed of such layers is termed a convolutional neural network (CNN) [43].

A 1D convolution operation maps an input vector $\mathbf{x}$ to an output vector $\mathbf{z}$ by computing weighted sums of local input regions. Each element z_i of the output results from applying identical weights, collectively known as the convolution kernel or filter, to a corresponding neighbourhood in the input. The kernel size defines the spatial extent of this neighbourhood. For a kernel of size three, the operation is expressed as:

$$z_i = \omega_1 x_{i-1} + \omega_2 x_i + \omega_3 x_{i+1}, \tag{3.13}$$

where $\boldsymbol{\omega} = [\omega_1, \omega_2, \omega_3]^T$ represents the kernel weights, the trainable parameters.

Several parameters modify convolutional operations:

Padding refers to the addition of values (typically zeros) around the input boundaries to control the output dimensions. Zero padding preserves the spatial dimensions of the input by adding zeros around its borders, ensuring border elements receive equal treatment in the convolution.

Stride is the number of pixels the kernel moves at each step. A stride of 1 preserves spatial resolution, while a stride of 2 halves the output's spatial dimensions.

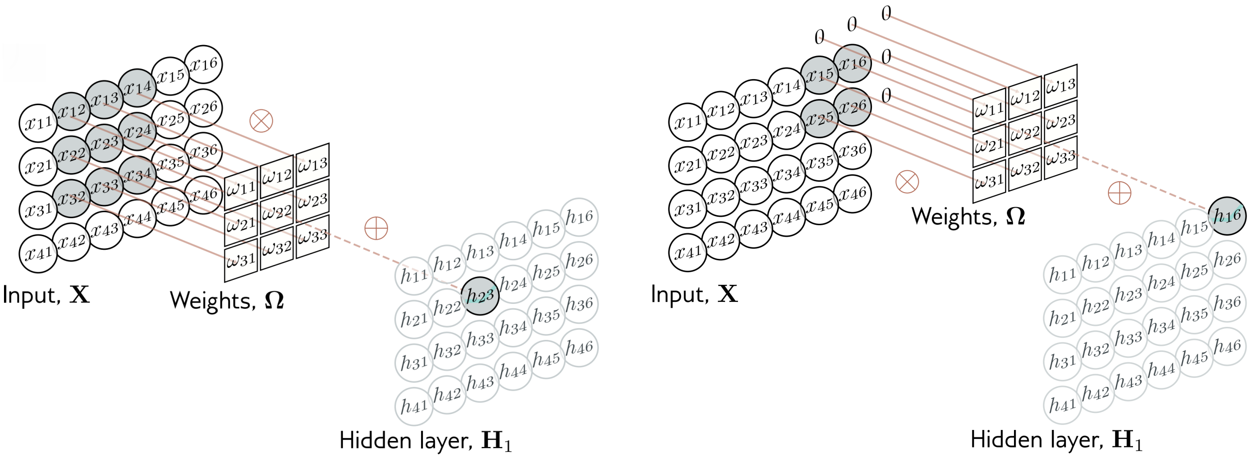
Kernel size defines the receptive field of the convolution. Larger kernels integrate information from wider regions but require more parameters. Odd-sized kernels (such as 3, 5, 7, etc.) are typically used to maintain symmetry around the central position.

Dilation introduces gaps between kernel elements, effectively expanding the receptive field without increasing the parameter count. A dilation rate of d inserts $(d-1)$ zeros between consecutive kernel elements, allowing the network to capture broader spatial dependencies with the same number of parameters.

Figure 3.2 illustrates a convolutional network for 2D inputs, such as image data. In such case, the convolution kernel is now a 2D object, and the hidden units h_{ij} in the case of a 3x3 kernel $\Omega \in \mathbb{R}^{3x3}$ are computed as:

$$h_{ij} \;=\; a\left[\beta + \sum_{m=1}^{3}\sum_{n=1}^{3} \omega_{mn} x_{i+m-2, j+n-2}\right]. \tag{3.14}$$

The kernel is translated both horizontally and vertically across the 2D input to create an output at each position. The most common activation function in modern neural networks is the rectified linear unit or ReLU 3.3:



(a) (b)

Figure 3.2: 2D convolutional layer, each output h_{ij} is a weighted sum of the 3x3 nearest inputs, with a bias and passes through an activation function. (a) The output h_{23} is a weighted sum of the nine positions from x_{12} to x_{34}. (b) With zero-padding, positions beyond the image's edge are considered to be zero. [58]

$$\text{ReLU}[z] = \begin{cases} 0, & \text{if } z < (\\ z, & \text{if } z \geq (\end{cases} \quad (3.15)$$

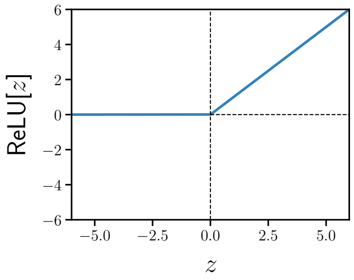
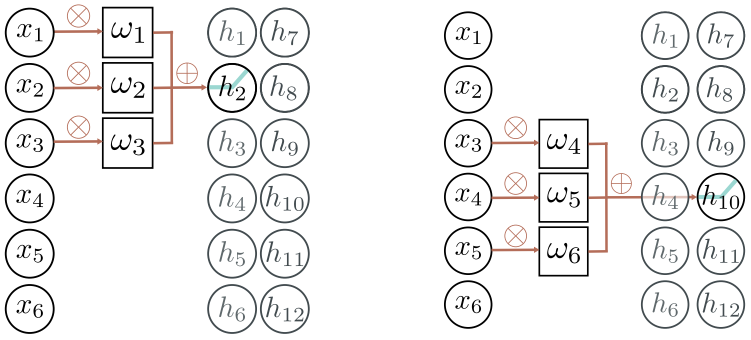


Figure 3.3: The ReLU activation function returns 0 for negative inputs while preserving positive inputs unchanged. This piecewise linearity enables efficient gradient computation while introducing the nonlinearity necessary for modelling complex functions.

3.4.1 Convolutional layers

A single convolution operation may lose information as it essentially averages neighbouring input values and eliminates negative responses through the ReLU activation as represented in Figure 3.4. To preserve and extract diverse features from the input, convolutional layers typically compute multiple parallel convolutions, each with different learned parameters. Each convolution creates a set of new hidden units known as a feature map or channel.



(a) (b)

Figure 3.4: Illustration of two separate convolution operations applied to a one-dimensional input signal **x** using a kernel of size 3. (a) The first convolution generates hidden units h_1 through h_6, which constitute the first channel. (b) The second convolution produces hidden units h_7 through h_{12}, forming a second, distinct channel. Adapted from [58].

These hidden units represent learned intermediate representations of the data. For example, when processing 2D data (such as images or spatial patterns), early convolutional layers may develop hidden units that respond to basic visual elements, like edges, corners, or textures, at different orientations.

Subsequent layers combine these elementary features to detect progressively more complex and abstract patterns.

The number of feature maps (i.e., the number of parallel convolutions) in each layer is a critical architectural hyperparameter that determines the model's capacity to represent complex patterns. For 2D data applications, a typical convolutional layer typically contains anywhere from 8 to 512 feature maps, or even more, with deeper layers generally having more channels to capture increasingly sophisticated representations.

3.4.2 Downsampling and upsampling

In convolutional neural networks, downsampling operations are essential for increasing the receptive field and reducing spatial dimensions, thereby enabling the network to capture larger-scale features. However, when the desired output is also an image (as in segmentation or generative tasks), a corresponding upsampling operation becomes necessary to restore the original spatial dimensions.

Several downsampling methods exist for 2D representations. In this work, 2D max pooling is employed due to its ability to introduce translation invariance. In max pooling, each output pixel represents the maximum value from the corresponding 2×2 input block, as illustrated in Figure 3.5. This operation not only reduces spatial dimensions but also helps in extracting dominant features while discarding less relevant information.

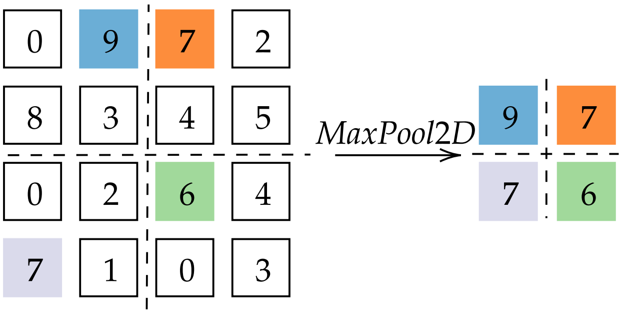


Figure 3.5: Illustration of 2D max pooling operation. Each output value is computed as the maximum of the corresponding 2×2 input block, resulting in a downsampled representation with half the spatial dimensions.

Transposed convolution is employed for upsampling, which effectively reverses the downsampling process by doubling spatial dimensions. With a kernel size of 2 and a stride of 2, it effectively doubles the spatial resolution while maintaining learnable parameters that adapt during training to reconstruct relevant features.

3.4.3 Batch normalization

Batch normalisation (BatchNorm) shifts and rescales according to the mean and variance estimated on the batch. [32]. Typically inserted between the convolution operation and the activation function (e.g., ReLU), BatchNorm introduces just two learnable parameters per feature channel: a scaling factor γ and a shift parameter β.

This technique yields multiple benefits: it smooths the loss landscape, permits larger learning rates, accelerates training convergence, acts as a regularizer by introducing beneficial noise, and stabilises forward propagation by mitigating internal covariate shift [32]. These advantages have made BatchNorm a standard component in modern deep neural architectures, significantly improving both training efficiency and generalisation performance.

3.5 U-Net architecture and the semantic segmentation task

The U-Net architecture was originally introduced by Ronneberger et al. [63] for biomedical image segmentation. It is based on a fully convolutional network and follows a symmetric encoder-decoder structure. The architecture is designed to perform precise pixel-wise classification while preserving both global context and fine-grained spatial information.

In its original form (see Figure 3.6), U-Net consists of two main paths:

- **Contracting path (encoder)**: This part of the network extracts increasingly abstract features using repeated applications of convolutional layers (with ReLU activation), followed by 2 × 2 max-pooling operations that reduce the spatial resolution. At each downsampling stage, the number of feature channels doubles.

- **Expanding path (decoder)**: This path gradually reconstructs the spatial resolution using up-convolutions (or transposed convolutions), reducing the number of feature channels and concatenating corresponding high-resolution feature maps from the encoder via skip connections. These skip connections, sometimes referred to as "copy and crop" in the original paper, enable the decoder to recover spatial details lost during downsampling.

The network ends with a 1 × 1 convolution that reduces the number of output channels to match the number of classes in the segmentation task, typically followed by a sigmoid activation function for binary segmentation or a softmax activation for multi-class segmentation.

Semantic Segmentation. Semantic segmentation involves assigning a class label to every pixel in an image, effectively transforming the task into a dense classification problem. The goal is to predict a map of class probabilities, one per pixel.

Formally, given an input image of dimensions $\text{Res} \times \text{Res} \times C_{\text{in}}$, where C_{in} is typically 1 for grayscale images or 3 for RGB, the U-Net outputs a tensor of size $\text{Res}' \times \text{Res}' \times N$, where N is the number of target classes and Res' is the spatial resolution of the output. In many implementations, $\text{Res}' = \text{Res}$, but this is not strictly required and depends on the padding strategy and architecture details. Each pixel in the output corresponds to a probability distribution over all possible classes. The final segmentation mask is obtained by selecting the class with the highest predicted probability for each pixel.

3.5.1 Skip connections in U-Net

A key feature of the U-Net architecture is its use of skip connections, which link corresponding layers in the encoder and decoder. These connections address the loss of spatial detail caused by downsampling operations in the encoder (e.g., max-pooling), which, while capturing high-level features, reduce localisation accuracy.

To mitigate this, U-Net concatenates encoder feature maps with the decoder's upsampled outputs along the channel dimension:

$$\mathbf{y} = \text{Concat}(\mathbf{x}_{\text{encoder}}, \mathbf{x}_{\text{decoder}}) \tag{3.16}$$

This fusion merges deep semantic information with fine details from early layers, enhancing segmentation map accuracy while preserving spatial context.

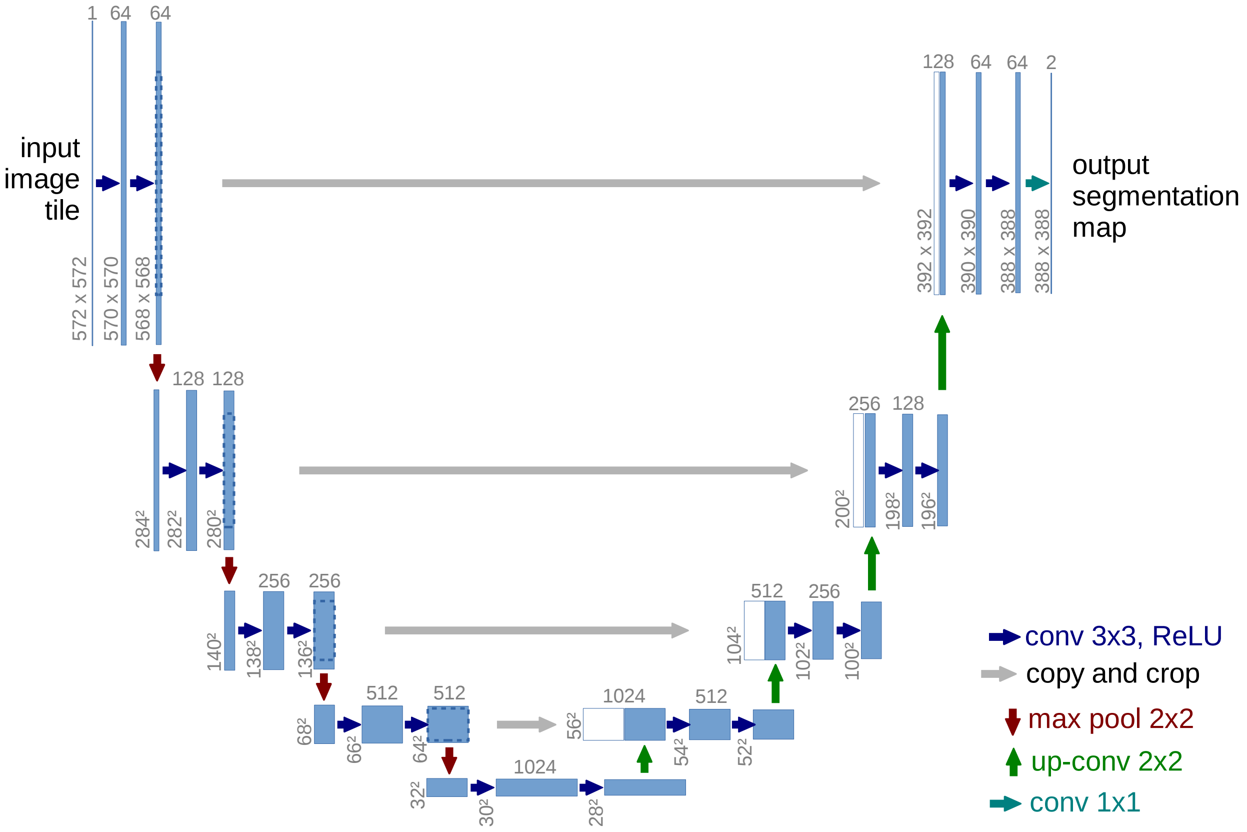


Figure 3.6: U-Net architecture example, shown for a case where the lowest resolution feature map is of size 32×32 pixels. Each blue box represents a multi-channel feature map, with the number of channels indicated at the top of each box. The spatial resolution (height × width) is displayed at the lower-left corner of each box. White boxes indicate feature maps that are copied and concatenated during the expansive path. Adapted from [63].

The architecture illustrated above comprises several critical model hyperparameters that collectively shape the network's performance characteristics:

- Depth: The number of downsampling/upsampling operations defining the contracting and expansive paths.
- Number of convolutions: The number of convolutional operations performed at each resolution level.
- Initial number of filters: The number of channels in the first layer, typically doubled after each downsampling operation.
- Image resolution, Batch size, and learning rate.

These hyperparameters are not static values but tunable parameters whose selection can markedly influence the accuracy and robustness of the network's predictions. Determining their optimal values typically requires systematic hyperparameter optimisation.

The performance of deep learning models depends significantly on these carefully tuned hyperparameters, which must be adapted to the specific characteristics of the dataset, making optimisation a fundamental step in model development [8, 44, 69].

Several hyperparameter optimisation strategies have been explored in the literature, each offering unique trade-offs. Grid search represents the most straightforward approach, involving an exhaustive evaluation of all possible combinations within a manually specified subset of the hyperparameter space.

Random search offers an alternative by sampling hyperparameter configurations stochastically from predefined distributions. Research by Bergstra and Bengio [8] has demonstrated that random search frequently outperforms grid search in efficiency, particularly when only a subset of hyperparameters significantly affects model performance. Nevertheless, random search exhibits a fundamental limitation: it retains no memory of previous evaluations and may, therefore, redundantly explore unpromising regions of the hyperparameter space.

While identifying optimal hyperparameters remains important, the optimisation process demands substantial computational resources and time, as each hyperparameter combination necessitates training and evaluating a complete model. Furthermore, this process demonstrates hardware dependence, as increased model capacity typically corresponds to greater computational requirements.

Given these constraints, the present research does not aim to find the absolute optimal hyperparameters for the problem but instead seeks a balanced set of hyperparameters that provide good efficiency while maintaining a moderate model capacity. Rather than aiming for exhaustive optimisation, this approach prioritises efficiency, acknowledging the computational limitations commonly encountered in deep learning workflows.

The hyperparameter optimisation for this research is structured in two stages and is depicted in Figure 3.7. In the first stage, the focal loss parameters α and γ are optimised via a grid search, while the architectural hyperparameters remain fixed. Direct comparison of raw focal loss values (see Eqn. 3.2.1) across different (α, γ) combinations does not reflect segmentation performance. Instead, three independent metrics are employed:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{F1} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3.17}$$

Here, TP (true positives) denotes the correctly predicted unstable branch pixels, FP (false positives) are pixels incorrectly predicted as unstable, and FN (false negatives) are truly unstable pixels that the model failed to predict.

The classification threshold τ is treated as a tunable parameter: for each (α, γ) pair, $\tau \in [0, 1]$ is swept to maximize the chosen metric. Optimal focal-loss hyperparameters (α^*, γ^*) and threshold τ^* correspond to the highest F1 score.

In the second stage, with α^*, γ^*, and τ^* fixed, architectural hyperparameters are explored via random search. This two-stage approach significantly reduces the total number of model evaluations while acknowledging that potential interactions between loss and architecture may remain unexplored.

Note that the focal loss parameters can exhibit sensitivity to dataset characteristics; details of dataset generation and preprocessing are provided in the following chapter.

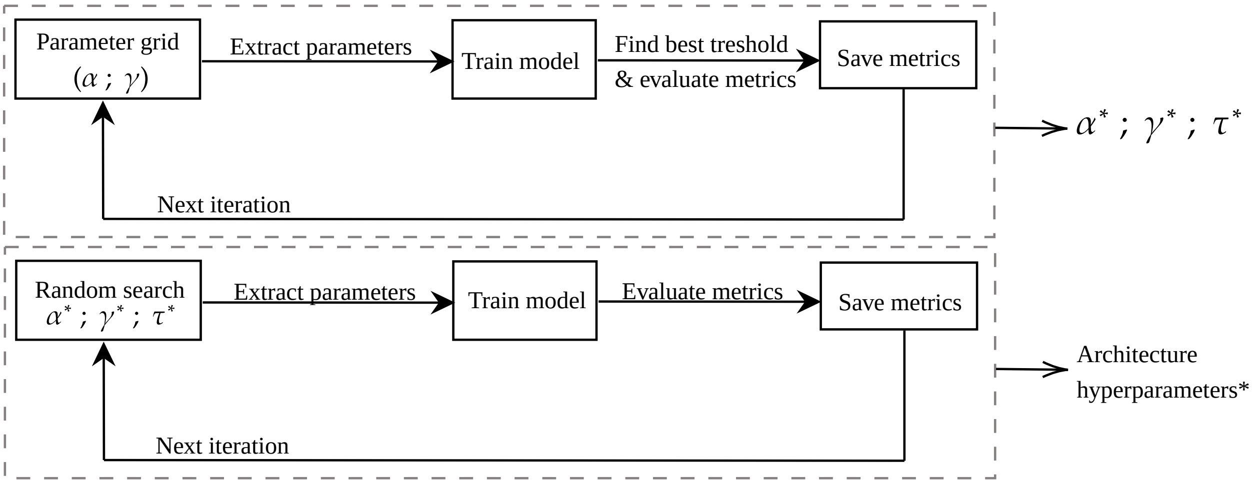


Figure 3.7: Two-stage hyperparameter optimisation procedure.

3.5.2 Implementation details

The implementation utilised Python and the PyTorch deep learning framework. The modular model architecture allowed for easy modifications and hyperparameter tuning. The training was performed on the Alan GPU cluster at the University of Liège, leveraging multiple GPUs and CPUs for efficient computation.

3.6 Chapter conclusion and discussion

Chapter conclusion

This chapter introduced the fundamental concepts of machine learning with an emphasis on those directly relevant to this research.

The principles of supervised learning were presented, highlighting the central idea of constructing a model that maps inputs to outputs via trainable parameters [58]. This led to the formulation of the learning process as an optimisation problem governed by a loss function that quantifies the mismatch between predictions and the ground truth. The case of binary classification was examined in detail, including the definition of the associated loss functions, binary cross-entropy and focal loss [46].

Next, convolutional layers were discussed as the foundational building blocks of the architecture used in this research [43]. Their key components and properties, such as weight sharing, receptive fields, and translation equivariance, were described in detail.

Finally, the U-Net architecture [63] and its implementation for semantic segmentation tasks were presented, emphasising the importance of the encoder-decoder structure and the role of skip connections in preserving spatial information. Additionally, the two-stage hyperparameter optimisation strategy employed in this research was detailed.

Throughout this chapter, the connection between machine learning models and the input-output pairs forming the dataset was established. The next chapter will focus on generating the datasets required for training the model.

CHAPTER 4

Generation of the datasets

The previous chapter introduced supervised learning and how a trained model learns to map the input $\mathbf{x}$ to the output $\mathbf{y}$ by minimising a loss function, resulting in the optimised parameters $\boldsymbol{\phi}$ of the system. The training part of the model utilises a training dataset comprised of I input-output pairs $\{\mathbf{x}_i, \mathbf{y}_I\}_{i=1}^{I}$. This chapter introduces the concepts behind generating these input-output pairs and emphasises the importance of constructing a comprehensive and reliable dataset.

4.1 Size of the dataset

In many machine learning applications, research often focuses on finding the best models that perform optimally on already existing datasets. Consequently, several massive datasets comprising an enormous number of labelled examples have been developed, including the MNIST dataset for handwritten digit recognition, ImageNet for object classification in images, and numerous datasets in the biomedical domain. The size of the dataset plays a crucial role in the performance of predictive models; multiple studies have demonstrated that for many deep learning tasks, performance continues to improve logarithmically as dataset size increases [74, 79, 86]. This relationship underscores a fundamental principle in machine learning: without high-quality data in sufficient quantities, even the most sophisticated models cannot perform well.

For this specific research on FRCs, no existing datasets are available. Therefore, creating a reliable and comprehensive dataset represents an essential component of this research endeavour.

4.2 Formulation of the input-output pairs

The first step in dataset generation is defining appropriate input-output pairs. The ML model aims to predict unstable branches of a system, given its stable branches. The system inputs are the stable branches obtainable through swept-up and swept-down sine testing, as explained in Section 2.2.

The system outputs vary based on prediction targets. While ideally, a single model would predict all unstable branches (fundamental resonance, superharmonics, and subharmonics), practical constraints necessitate separate models for each type of resonance. This approach offers several advantages:

- It addresses resolution constraints, as superharmonics appear at much smaller amplitudes than fundamental resonances and subharmonics, which would otherwise require impractically high-resolution images for CNN processing.

- It enables targeted testing protocols. For example, superharmonic prediction requires swept-sine testing only at low frequencies, allowing faster testing over smaller frequency ranges with higher forcing without risking structural damage.
- It simplifies the learning problem by requiring each model to learn patterns for only one type of resonance.

This approach does introduce complexity in model development, as each model may require different optimal hyperparameters, potentially necessitating separate hyperparameter searches.

4.3 Dataset generation schemes

Having established the framework for input-output pairs, a significant number of input-output pairs need to be generated. As explained above, the performance continues to improve logarithmically as the dataset size increases. Furthermore, sampling the parameter space of the nonlinear oscillators is crucial; the dataset must contain input-output pairs representing a wide range of responses to ensure that the models are adaptable to many types of structures. The diagram 4.1 illustrates how the different datasets are generated, and each step of the schema is explained below

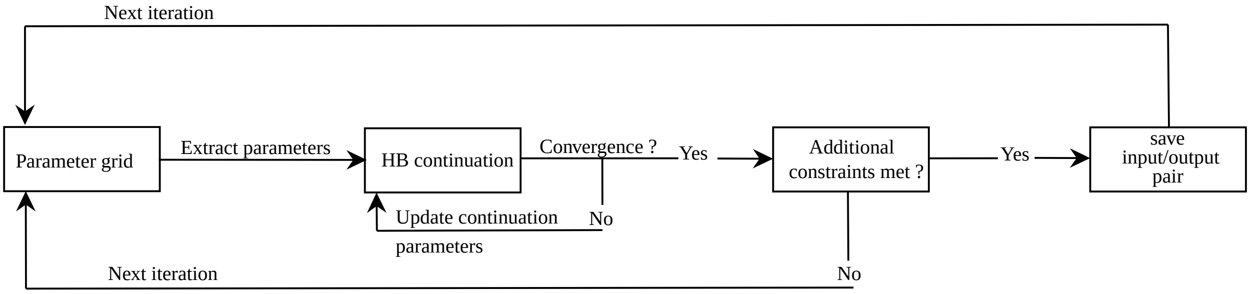


Figure 4.1: Dataset generation schema.

4.3.1 Parameter grid

As explained in Section 2.1, the equation of motion of the Duffing oscillator can be nondimensionalised, making this system dependent on two parameters, the damping ratio ζ, and the dimensionless force $\bar{F}$. The strategy behind generating the dataset is to simulate a large number of systems by varying these two parameters, and for each combination of these parameters, compute the FRC. The Helmholtz-Duffing and piecewise linear stiffness oscillators each introduce an additional parameter (β and κ, respectively).

An optimal parameter space must be:

1. Sufficiently large to capture diverse system responses
2. Adequately dense for proper sampling
3. Computationally feasible
4. Physically realistic (moderate forcing amplitudes and damping rations)

The selection of realistic parameter ranges is crucial; extremely low damping ratios ($< 0.5\%$) or excessive forcing amplitudes do not accurately represent physical systems, as actual structures would fail under such conditions.

4.3.2 HB continuation mechanisms

The HB continuation method implemented in this study employs robust numerical strategies to ensure accurate computation of FRCs. The HB pseudo-arclength continuation algorithm, originally proposed by Keller (1977) [5], enables the continuation of solution families beyond fold points.

This approach follows a predictor-corrector framework that systematically advances along the solution path. The pseudo-arclength scheme is coupled with a Newton-based solver to resolve the set of $2N_H + 1$ nonlinear algebraic equations arising from the HB formulation, with an additional arclength constraint that ensures convergence near critical points.

Adaptive step size A critical component for maintaining both computational efficiency and robustness in pseudo-arclength continuation is the adaptive control of the step size Δh_i, particularly in regions of high curvature or near bifurcation points, where Newton-type corrector steps may fail to converge. As illustrated in Figure 4.2, constant step sizes either waste computational resources in linear regions or fail near folds, where the solution path sharply turns, and standard corrector methods may diverge.

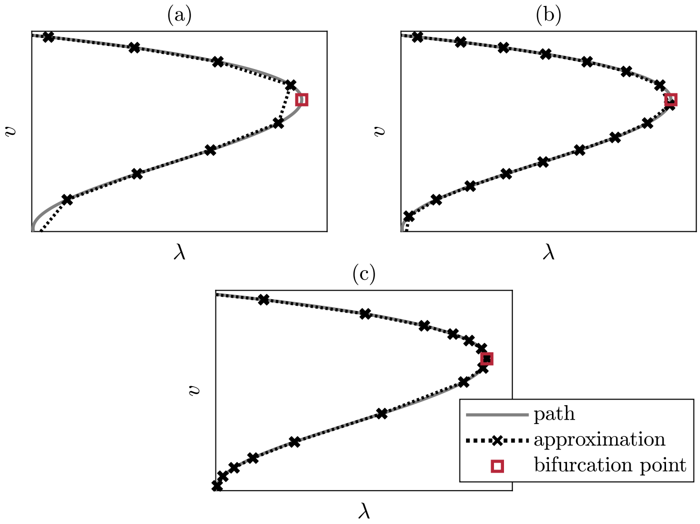


Figure 4.2: (a) Coarse constant step size, (b) finer constant step size, (c) variable step size. [41]

To address the difficulties associated with constant time steps, the step size is typically multiplied by a scaling factor ζ_i at each iteration:

$$\Delta h_{i+1} = \zeta_i \Delta h_i, \tag{4.1}$$

where Δh_{i+1} and Δh_i denote the step sizes at iterations $i+1$ and i, respectively. In practice, the step size is constrained by upper and lower bounds h_{max} and h_{min}. Additionally, maximum and minimum adaptation factors ζ_{max} and ζ_{min} are employed.

A widely used method for step size control adapts the step size based on the number of iteration steps N_i required during the Newton-Raphson method. Most approaches aim to achieve an optimal number of iteration steps N_{opt} for each solution point. Choosing a small N_{opt} demands fast convergence of the solver and results in smaller step sizes Δs_i, while a larger value allows for more iterations and enables larger step sizes. The ratio N_{opt}/N_i is commonly used to evaluate the scaling factor ζ_i:

$$\zeta_i(N_i) = \left(\frac{N_{\text{opt}}}{N_i}\right)^{\beta} \tag{4.2}$$

Seydel [66] uses $\beta = 1$, although other choices are possible, such as $\beta = 1/2$ as proposed by Ramm [59]. Alternative strategies involve adjusting step sizes based on the path's curvature.

Adaptive step size control is particularly crucial in this application, as it enables larger steps in well-behaved regions while automatically reducing step sizes when approaching challenging areas, such as fold bifurcations. This dynamic adjustment significantly accelerates the overall FRC computation without compromising the accuracy or stability of convergence.

In this work, the corrector step is implemented using a damped Newton method rather than a classical full-step Newton iteration. Damping improves convergence robustness in the presence of strong nonlinearities and ill-conditioned Jacobians, especially near folds and sharp transitions in the frequency response curve, where standard Newton methods often fail to converge [13].

Suppose convergence fails (typically due to exceeding the maximum number of Newton iterations or a stagnating residual). In that case, the step size is reduced by a fixed contraction factor $\zeta_{\text{fail}} < 1$, and the corrector step is retried from the previous solution point. This reduction and retry mechanism is iteratively applied until convergence is achieved or the step size falls below a predefined minimum h_{min}, at which point continuation is terminated.

The majority of the HB continuation implementation, including the computation of isolated responses, was adapted from [25]. The pseudo-arclength harmonic balance continuation approach with adaptive step size proved sufficient in terms of robustness and computational efficiency for the systems considered. Nonetheless, some advanced functionalities, such as branch switching, were not integrated due to the complexity of implementing them in a numerically stable and generalisable way.

Several specialised toolboxes are well-known for their robustness and widespread adoption in numerical continuation problems. These include established packages such as `MATCONT` [20] and `COCO` [2], which offer a wide range of built-in capabilities, including advanced bifurcation tracking and branch switching. Additionally, domain-specific frameworks such as `NI2D` [52] and `NLVib` [40] are tailored to nonlinear vibration problems, providing highly optimised routines for frequency response curve analysis.

4.3.3 Input/output representation for CNN

Following the continuation procedure, pairs of points (ω, A) are obtained, representing the FRC of a given oscillator. From these data, input-output pairs are constructed for training. The input consists of the stable branches of the FRC, while the output corresponds to the unstable branches specific to the model or the complete resonance in the case of the subharmonics.

Furthermore, the (ω, A) arrays must be converted into a grid-based representation compatible with convolutional neural networks. In particular, for U-Net architectures, the input dimensions must be multiples of 32 to accommodate the successive downsampling and upsampling operations.

The problem is formulated as binary classification at the pixel level:

1. Each pixel is assigned either 1 (curve/foreground) or 0 (background)
2. When visualised, these correspond to white and black pixels, respectively
3. The data is structured as H×W×1 tensors with binary values

Standard CNN implementations typically require an initial preprocessing step to convert images into tensors. However, direct tensor generation from FRC data offers significant advantages: reduced storage requirements compared to image formats, improved computational efficiency by eliminating the image-to-tensor conversion step, and an optimised processing pipeline for generating large-scale datasets.

For this direct tensor generation, Bresenham's line algorithm is implemented to create the grid representation. Bresenham's algorithm efficiently renders continuous FRC data onto discrete pixel grids using only integer operations. For a line segment between points (x_0, y_0) and (x_1, y_1) as illustrated in Figure 4.3, the algorithm determines the primary direction of movement (x or y axis), increments along the primary axis at each step, maintains an error accumulator to determine when to increment the secondary axis, and selects pixels that most closely approximate the analytical line. This approach minimises computational cost while preserving the visual accuracy of the curves, effectively converting continuous FRC data into the binary grid representation required for CNN processing.

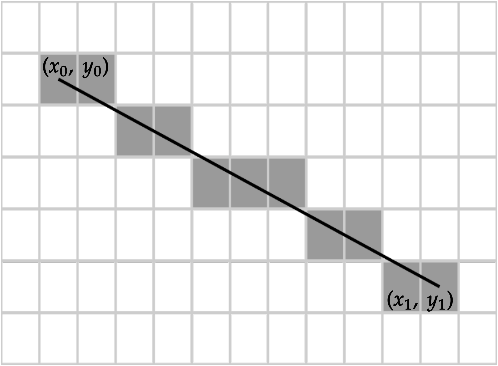


Figure 4.3: Diagram representing the Bresenham's line algorithm applied to a line segment between points (x_0, y_0) and (x_1, y_1). The grey pixels correspond to the foreground/curve pixels, and the white pixels correspond to the background pixels.

Alternative methods, such as Xiaolin Wu's line algorithm, implement anti-aliasing and could potentially produce smoother curve representations by assigning intermediate intensity values to partially covered pixels. However, such approaches would transform the task from binary classification to regression as the pixel values would lie in the continuous range $[0, 1]$ rather than being strictly binary. The binary representation adopted in this research strikes a balance between methodological simplicity and sufficient accuracy for the prediction task.

4.3.4 Additional conditions and image construction considerations

A set of additional criteria is imposed during the dataset generation to ensure physical relevance and nonlinear dynamical richness. First and foremost, the objective of this work is to study systems exhibiting pronounced nonlinear behaviour. Therefore, only samples whose FRCs contain at least one unstable branch are retained. The continuation process tracks the entire frequency response, and any samples lacking unstable branches are discarded. This filtering guarantees that the database is exclusively composed of systems demonstrating nonlinear behaviour.

In addition, an upper limit is imposed on the maximum frequency shift of the resonance. Specifically, the fundamental resonance frequency is constrained to remain within 500% of the undamped linear natural frequency. This choice is both physically and experimentally motivated. In practice, such large frequency shifts are rarely observed in experimental setups, especially when high forcing amplitudes and low damping are required to induce them. Including cases beyond this range could introduce unrealistic or unphysical data into the model, potentially compromising its applicability to real-world systems. Another motivation for limiting the maximum resonance shift lies in the design choice of using a fixed frequency window across all samples. If samples exhibiting large resonance shifts were retained, the frequency window would need to be expanded accordingly to ensure that the entire resonance is captured. Although such samples would be well-captured, as their resonance spans a substantial portion of the image, samples with more moderate shifts would appear compressed into a narrow horizontal region, reducing their representational resolution.

By restricting the maximum frequency shift, the selected frequency window achieves a better compromise: it is wide enough to fully represent the resonance for all included samples while maintaining high resolution for those with smaller shifts. This strategy ensures a uniform and informative image representation across the dataset, thereby enhancing CNN's capacity to generalise across different dynamical regimes.

4.4 Image representation strategy

Choice of frequency and amplitude windows. Each FRC is transformed into a 2D binary image where the horizontal axis represents the dimensionless frequency (normalised by the linear natural frequency), and the vertical axis corresponds to the amplitude. To enable CNN to learn spatial patterns across different samples effectively, it is crucial to adopt a consistent and physically meaningful image representation.

The horizontal frequency window is fixed across the entire dataset, ranging from 0 to 6, i.e., up to 600% of the linear resonance frequency. By ensuring that the fundamental resonance always appears within the same region of the image, the network can learn the typical shapes and structures of unstable branches more effectively.

A key aspect of this decision is the need to encompass the entire fundamental resonance within this frequency span. The location of the resonance peak varies significantly depending on system parameters such as damping and forcing amplitude. In particular, systems with low damping and strong excitation exhibit large resonance shifts.

On the other hand, extending the frequency window too far would reduce the spatial resolution of critical features in the image, as the image size is finite. For instance, unstable branches, often small and localised, would be under-sampled and poorly represented if the window were unnecessarily wide. The fixed limit of $6\,\omega_0$ provides a balanced trade-off: it ensures that the fundamental and possible superharmonic features are fully captured without overly sacrificing resolution.

Moreover, post-resonance behaviour is typically less informative. After the resonance peak, the amplitude tends to decay towards zero across all samples, offering little discriminative information. Therefore, allocating excessive resolution to this region is inefficient.

Amplitude axis standardization Unlike the frequency axis, the vertical (amplitude) axis is defined individually for each sample. The maximum amplitude is computed for each response, and the vertical axis is scaled accordingly. This adaptive normalisation ensures that the entire curve is visible in the image, regardless of the magnitude of the response. A global amplitude window would introduce excessive white space due to the high variability in peak amplitudes across samples. Figure 4.4 illustrates the resulting image representations for the prediction of the fundamental resonance of a Duffing oscillator.

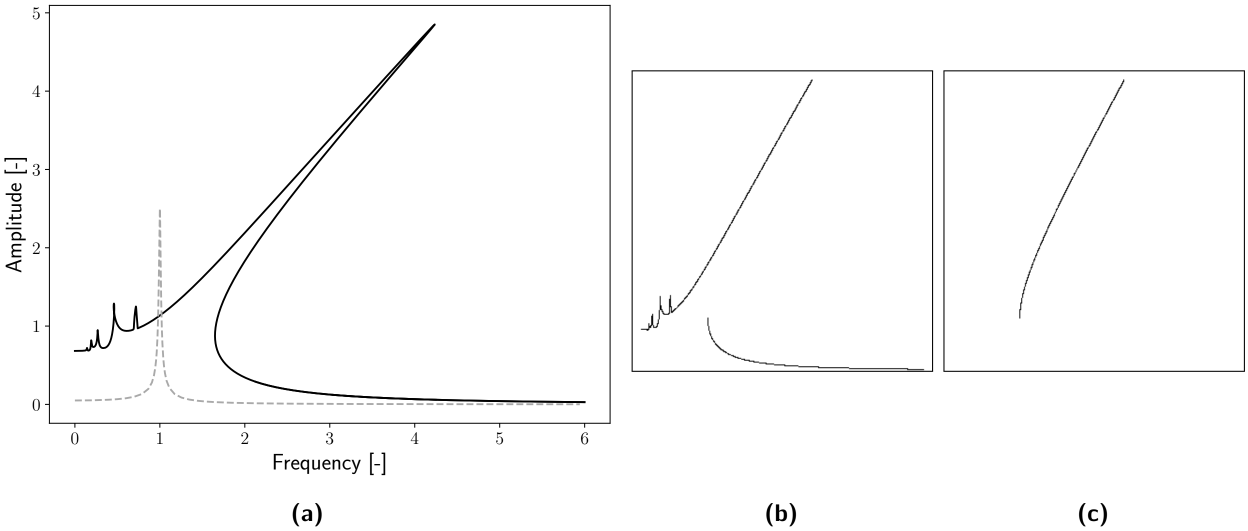


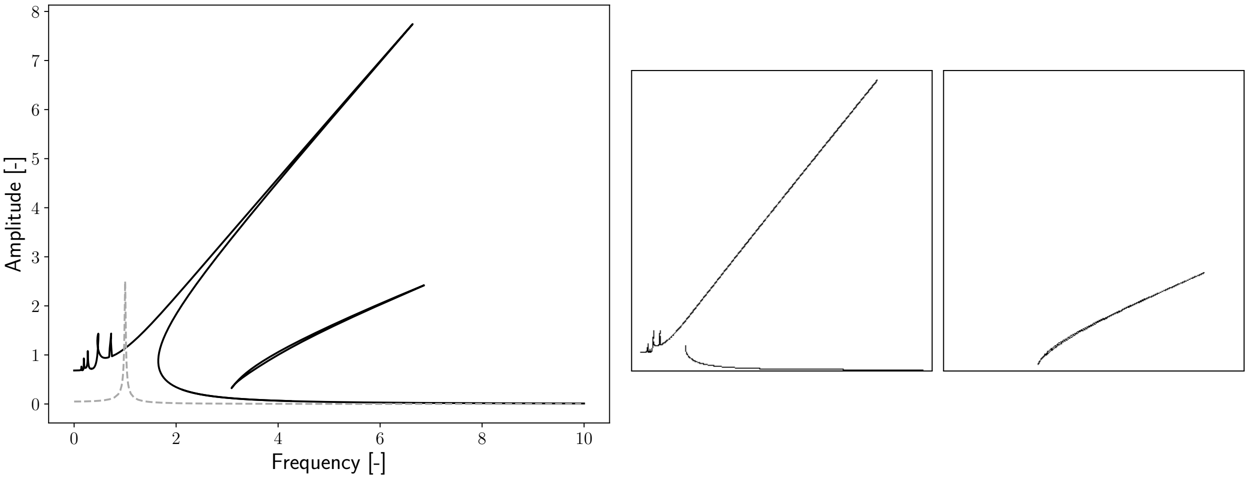
Figure 4.4: Example of the input/output image encoding used for the prediction of the fundamental resonance unstable branch. (a) displays the complete FRC, with the linear reference in gray. (b) and (c) show the corresponding 320×320 binary input and output images: (b) is constructed from the stable parts of the curve, and (c) represents the unstable branch of the fundamental resonance.

Experimental alignment: The proposed image encoding strategy is particularly suitable for integration with experimental data, as the parameters required to reproduce the normalisation procedure can be readily extracted from standard measurement techniques. In particular, a low-amplitude sine-sweep excitation enables the approximation of the natural frequency, which serves as the normalisation reference for the frequency axis. This allows experimental responses to be embedded in the same dimensionless frequency range as the simulated samples.

Regarding the amplitude normalisation, the vertical scaling is determined for each sample by identifying the maximum and minimum amplitudes within the selected frequency window. These extrema are directly observable in typical sine-sweep tests. Specifically, the peak amplitude corresponds to the point where the system exhibits a jump-down, while the smallest amplitude within the window (often located at the frequency boundary) sets the lower bound. As both points are accessible during the frequency sweep, the full amplitude window required for normalisation can be experimentally obtained.

This alignment ensures that experimental data can be post-processed to produce input representations identical to those used for training on numerical data.

Applicability and Extensions. It is important to emphasise that the strategy described above is specific to the encoding of the fundamental resonance. A similar methodology is employed for encoding subharmonic isolas, with modifications to the frequency window to account for the distinct range over which these features occur (see Section 5.2.2). Figure 4.5 illustrates the conversion process from raw FRC data (Figure 4.5a) to binary image representations used as CNN input (Figure 4.5b) and output targets (Figure 4.5c).



(a) **(b)** **(c)**

Figure 4.5: Illustration of the image generation process from a FRC. (a) shows the complete FRC, with the grey line indicating the equivalent linear response; (b) and (c) respectively display the 320×320 binary input and output images, where (b) is constructed from the stable branches and (c) represents the subharmonic resonances.

A notable advantage of the proposed representation is its flexibility: the frequency and amplitude windows, as well as image resolution, can be easily adjusted to suit other types of nonlinear responses or experimental setups. While this method was adopted in the present work for its simplicity and efficiency, alternative encoding strategies may offer advantages for specific tasks and remain an open avenue for future research.

In contrast to the encoding strategy adopted for fundamental and subharmonic resonances, superharmonic responses are treated differently. Each superharmonic is framed and encoded independently. This approach is motivated by both practical and methodological considerations. First, it enables scalability: in this work, only the 3:1 and 2:1 superharmonics are considered. By isolating each harmonic in its own frame, the dataset can be constructed specifically around these cases, excluding others that fall outside the scope of the prediction task.

Encoding each superharmonic independently also facilitates higher resolution in both frequency and amplitude. Since the frequency range of interest around a given superharmonic is narrow, defining a local window ensures a fine frequency resolution. This results in better discretisation of the resonance features within the image representation.

However, a major limitation of this strategy arises in the definition of the amplitude axis. For the fundamental and subharmonic cases, the amplitude axis is defined individually for each sample based on the maximum measured amplitude. This is not directly applicable to superharmonics, as the point of maximum amplitude may lie on an unstable branch and, therefore, remain inaccessible in swept-sine experiments, unlike the fundamental resonance, where the peak response is generally stable.

To address this, the encoding strategy for superharmonics is based solely on measurable quantities from the swept-sine data. As illustrated in Figure 4.6, the encoding window is defined by the horizontal distance *d* between the folds and a vertical distance *L* determined by the highest and lowest amplitude values within the frequency range.

This encoding strategy was selected primarily for its simplicity and reliance on experimentally accessible measures, in particular, the fold points that bound the stable portion of the superharmonic response. Since these folds are directly observable in swept-sine data, they provide a robust and reproducible basis for defining the encoding window. While alternative schemes could potentially offer greater generality or flexibility, the chosen approach provides a pragmatic compromise between simplicity, experimental applicability, and consistency in dataset construction.

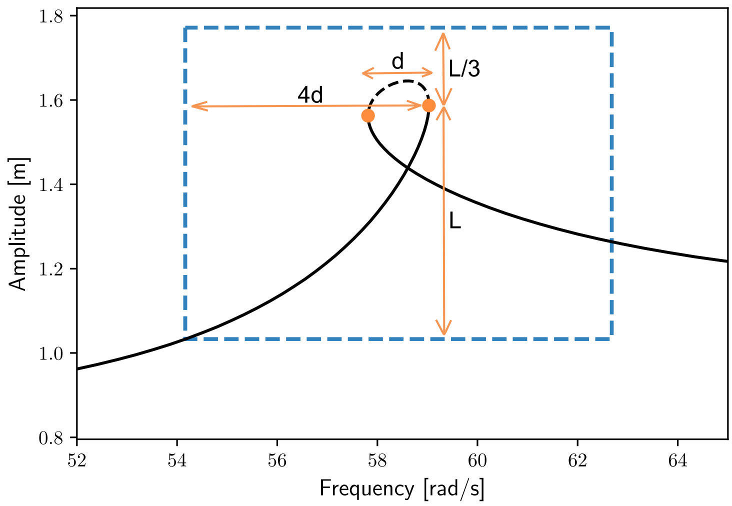


Figure 4.6: Illustration of the encoding strategy for superharmonic responses. The blue rectangle denotes the region of interest used to generate the image. The orange arrows illustrate the measurements used to construct the encoding frame around the superharmonic response.

Image resolution selection. The choice of image resolution is a key implementation consideration, especially in the context of convolutional architectures such as U-Net. The final resolution must satisfy architectural constraints namely that both width and height be multiples of 32 to accommodate successive pooling and upsampling operations. Beyond these constraints, the resolution significantly influences both the representational accuracy and the computational load.

In practice, the image resolution is primarily constrained by hardware limitations. Lower-resolution images are faster to generate and process and require less memory. However, they often lack the spatial fidelity necessary to capture finer features, particularly the superharmonic which tend to occupy narrow regions. If the resolution is too low, such features may be underrepresented or entirely lost, thereby limiting the model's predictive capabilities.

Conversely, increasing the resolution improves the clarity and fidelity of the binary curves, enabling a more accurate representation of complex behaviours. This is especially important for distinguishing between closely spaced branches. The trade-off is increased computational cost, both in terms of memory usage and training time, which may become prohibitive depending on available resources.

To balance these considerations, the resolution for this research was fixed at 320 × 320 pixels. This configuration offers sufficient spatial resolution to accurately depict all relevant dynamic features while remaining compatible with standard GPU hardware and allowing reasonable training times. It thus represents a pragmatic compromise between computational efficiency and physical accuracy.

This standardised window and resolution strategy ensures that the dataset is not only suitable for deep learning applications but also preserves the essential dynamical features necessary for the accurate prediction of unstable branches.

4.5 Chapter conclusion and discussion

Chapter conclusion

This chapter presented a comprehensive framework for generating datasets tailored to the supervised training of machine learning models aimed at predicting unstable branches in FRCs of nonlinear dynamical systems.

The construction of input-output pairs was based on separating stable and unstable segments from FRCs computed via HB continuation. Distinct models were adopted for each resonance type (fundamental, subharmonic, and superharmonic) to accommodate resolution limitations and reduce the complexity of the learning task.

A systematic parameter grid strategy was implemented to sample a wide range of physical behaviours. The HB continuation algorithm was enhanced with adaptive step-size control and robust corrector mechanisms to ensure numerical stability and computational efficiency, particularly near bifurcation points.

To enable direct processing by convolutional neural networks, FRC data were encoded as binary tensors using Bresenham's line algorithm. Each sample was mapped to a fixed 320×320 image format, with a standardised frequency axis and adaptive amplitude scaling to preserve structural features across varying dynamical regimes.

This pipeline delivers a scalable, automated, and physically grounded methodology for dataset generation. It balances accuracy and efficiency while ensuring compatibility with both numerical and experimental data, thereby providing a solid foundation for the subsequent machine learning tasks addressed in this thesis.

CHAPTER 5

Prediction of the resonances of a Duffing oscillator

In this chapter, the machine learning model introduced in Section 3 and the dataset generation strategy described in Section 4 are applied to the Duffing oscillator.

The chapter is structured into three main sections, each devoted to the prediction of a specific type of resonance observed in the Duffing oscillator: the fundamental resonance and the secondary resonances, namely subharmonics and superharmonics.

The general workflow involves generating a comprehensive dataset of FRCs, followed by the development and training of dedicated machine learning models tailored to each resonance type. The hyperparameter optimisation strategy, previously outlined in Section 3, is employed to ensure optimal model performance. The hyperparameters identified through this process are subsequently adopted throughout the remainder of the study.

This chapter begins with an in-depth analysis of the fundamental resonance, detailing the generation of the dataset, the partitioning strategy, and the model optimisation stages. The performance of the model is then assessed both qualitatively and quantitatively, emphasising its ability to generalise to previously unseen configurations of the Duffing oscillator.

5.1 Prediction of the fundamental resonance

The initial model examined is designed to forecast the unstable branch of the fundamental resonance within a Duffing oscillator. As explained in the preceding chapter, the machine learning model leverages input-output pairs, wherein the inputs consist of the stable branches of the FRCs, and the outputs are their respective unstable branches. Herein, the output specifically refers to the unstable branch associated with the fundamental resonance.

5.1.1 Dataset generation for the fundamental resonance

To generate a dataset with diverse samples, the algorithm described in Section 4 is implemented. The initial step involves defining the parameter grid, which involves selecting appropriate ranges for the damping ratio ζ and nondimensional force $\bar{F}$, as well as determining the number of discrete values within these ranges. The parameter ranges must reflect realistic conditions while encompassing a variety of responses to ensure model versatility across different Duffing oscillator configurations. Additionally, the sampling density within these ranges is crucial, as a sufficient number of samples must be included to represent the solution space adequately.

Another important parameter in the HB method is the harmonic truncation order N_H. While increasing N_H generally improves the accuracy of the periodic solution by capturing more harmonic content, it significantly increases the computational cost. This trade-off is illustrated in Figure 5.1, which shows the average computation time required to evaluate an FRC for varying values of N_H. The reported average computation time corresponds to the average time to compute an FRC dataset over a 10×10 parameter grid (see Table 5.1).

Although increasing N_H improves the fidelity of the solution, the marginal gain diminishes beyond a certain order due to the rapid decay of higher harmonic components. Although higher harmonics can influence lower-order components through nonlinear coupling, their overall contribution diminishes as the harmonic order increases. Furthermore, the computational complexity of the HB method combined with Newton–Raphson iteration scales as $\mathcal{O}(N_H^3)$.

Therefore, a balance must be struck between computational efficiency and accuracy, particularly in this work where the FRCs are rendered on a finite-resolution grid. As a result, only the dominant harmonic components and the first few superharmonics (e.g., 2:1, 3:1) are clearly visible, while higher-order contributions become visually indistinguishable and thus contribute little to the interpretability of the response curves. Based on convergence analysis (see Figure 5.1), a truncation order of $N_H = 7$ was selected as an appropriate compromise for constructing the fundamental resonance dataset.

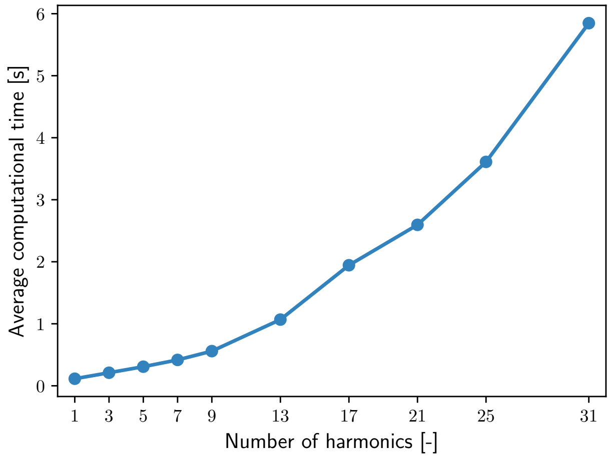


Figure 5.1: Average computational time per sample for evaluating a single FRC as a function of harmonic truncation order N_H over a 10×10 parameter grid (see Table 5.1).

The parameter space used to generate the fundamental resonance dataset, based on Equation 2.10, is summarised in Table 5.1.

Parameter	Range of values	Number of points
Damping ratio ζ [%]	$[0.5, 25]$	150
Dimensionless force $\bar{F}$ [$-$]	$[0.1, 3]$	150

Table 5.1: Parameter values used for generating the Duffing oscillator dataset for fundamental resonance prediction.

The resulting dataset is illustrated in Figure 5.2, where two prominent empty regions are visible. As explained in the previous chapter, additional constraints were imposed during the generation process: each FRC must include an unstable branch and exhibit a resonance shift of less than 500%. The empty region at high forcing and low damping corresponds to cases where the resonance shift exceeds this threshold, while the region at low forcing and high damping contains responses without unstable branches in the fundamental resonance and was therefore excluded.

Out of the $150^2 = 22{,}500$ attempted FRC continuations, only 13 failed to converge (0.06%), highlighting the robustness of the continuation algorithm. These rare failures are typically associated with numerical difficulties near folds. In such regions, the continuation algorithm, in cases of failure, reversed direction prematurely, thereby failing to compute the complete FRC. After applying the aforementioned constraints, the final dataset consists of slightly more than 18,000 valid samples.

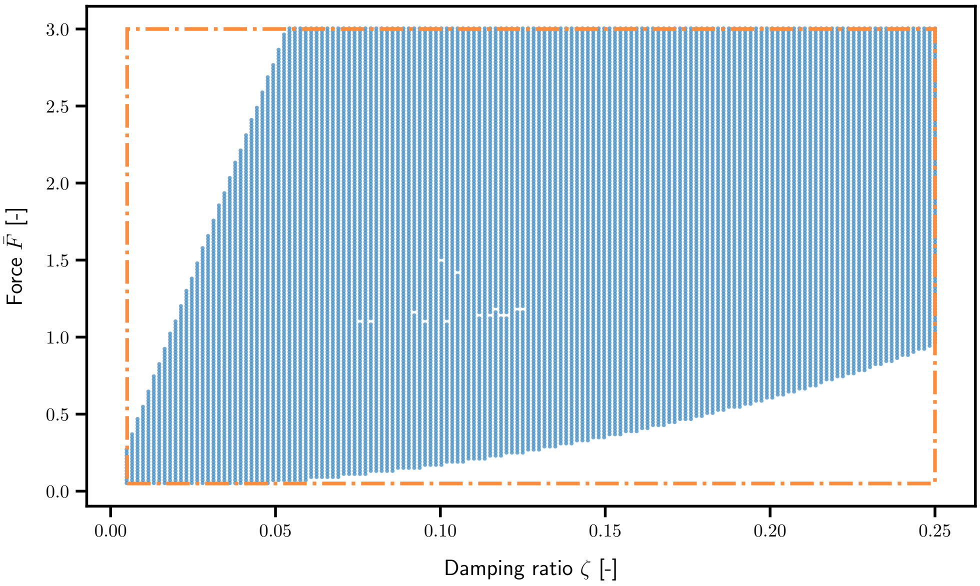


Figure 5.2: Representation of the parameter space of the dataset for the fundamental resonance of the Duffing oscillator. The blue dots represent the samples.

5.1.2 Dataset partitioning strategy

As discussed in Section 3, model training is a critical phase where the parameters of the model are learned. However, the hyperparameters of the system also significantly influence the model's generalisation capability [26]. Increasing a model's capacity does not necessarily improve its generalisation capability, even if it enables better prediction of training samples. This occurs because the training algorithm seeks to find model parameters $\boldsymbol{\phi}$ that minimise the loss function on the training data, but this optimisation does not guarantee performance on unseen data [9].

From a mechanical engineering perspective, a strong generalisation capability is essential, as it enables the model to accurately predict the responses of mechanical systems with properties different from those used during training. In the case of the Duffing oscillator, this means predicting behaviour across various damping ratios ζ and forcing amplitudes $\bar{F}$.

To evaluate generalisation capability, the dataset is typically divided into distinct subsets [30]. The training set, which comprises the majority of the data, is used for model parameter optimisation. The validation set, independent from the training set, serves to measure model performance and guide hyperparameter selection. While the training loss typically decreases monotonically with epochs, the validation loss often exhibits a minimum, beyond which overfitting occurs. The test set provides a final, unbiased evaluation of model performance after hyperparameter optimisation is complete.

The conventional dataset partition allocates approximately 80%, 10%, and 10% to training, validation, and test sets, respectively [42]. However, this research employs a different strategy, illustrated in Figure 5.3. The test set occupies the central region of the parameter space, while the training and validation sets are drawn from the remaining regions, with validation samples randomly selected from this space.

This partitioning strategy enables the assessment of different model capabilities. The validation set evaluates interpolation performance, as its samples are proximate to training samples in parameter space. Conversely, the test set, occupying a distinct region where training samples are absent, measures the model's extrapolation or generalisation performance and its ability to predict system behaviour for parameter combinations different from those encountered during training.

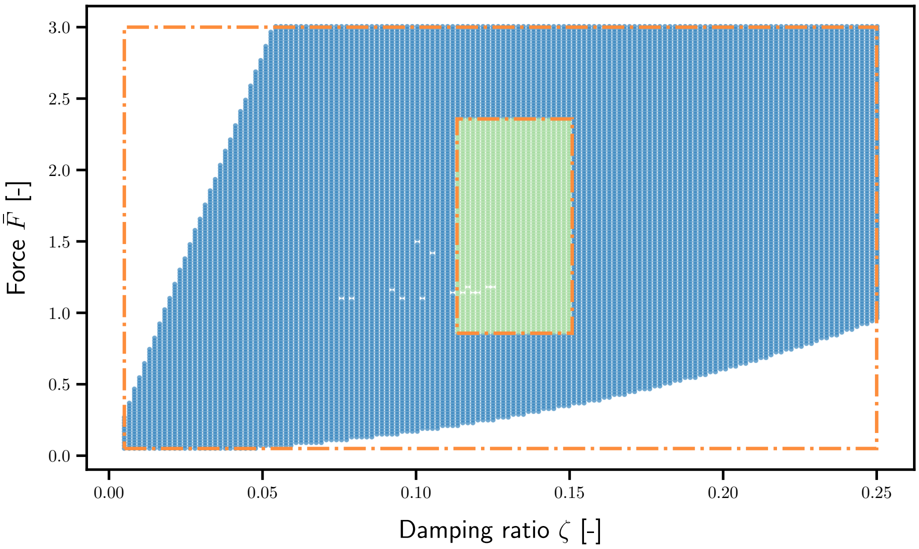


Figure 5.3: Illustration of the dataset partitioning strategy, showing the central test region (green dots), and surrounding training/validation regions in parameter space (blue dots).

5.1.3 Focal loss hyperparameter optimisation

The hyperparameter optimisation strategy described in Section 3.5 is applied based on the dataset partitioning scheme illustrated in Figure 5.3. The optimisation follows a two-stage approach: the first stage focuses on tuning the focal loss parameters α and γ while keeping the network architecture fixed, whereas the second stage, detailed later, involves optimising the architectural hyperparameters.

The goal of the first stage is to assess the impact of various (α, γ) combinations on the model performance, using the F1 score computed on both the validation and test sets as the evaluation metric. For each configuration, the classification threshold τ is optimised to maximise the F1 score. The F1 score evolution over training epochs for selected configurations is shown in Figure A.1, and a summary comparison of the results is provided in Table A.1. For reference, performances obtained with the standard binary cross-entropy loss are also reported.

	Validation set F1 score [%]	**Test set F1 score [%]**
Focal Loss ($\alpha = 0.8, \gamma = 3.0$)	83.48	83.61
BCE Loss	81.05	82.67

Table 5.2: Comparison of validation and test F1 scores using focal loss ($\alpha = 0.8$, $\gamma = 3.0$) and standard binary cross-entropy (BCE) loss.

The optimal configuration was found for $\alpha = 0.8$ and $\gamma = 3.0$. Nonetheless, the improvement achieved by using focal loss over the BCE loss is relatively modest, yielding approximately a 2% increase in the F1 score, as shown in Table 5.2. This suggests that, while focal loss helps mitigate class imbalance and emphasises harder-to-classify samples [46], the existing network architecture and training strategy already address these challenges to a significant extent.

5.1.4 Architectural hyperparameter optimisation

In contrast, the second stage of hyperparameter optimisation focusing on architectural parameters results in substantial improvements in performance. Specifically, when comparing the best architecture found through random search to the original U-Net configuration proposed in [63], an increase of approximately 10% in the F1 score is observed, as shown in Table 5.3. This highlights the critical importance of performing a dedicated hyperparameter optimisation rather than directly adopting standard architectures. The different model hyperparameters for the various runs performed during the random search are listed in Table A.2, and the evolution of the validation and test F1 scores over training epochs is depicted in Figure A.2.

	Validation set F1 score	**Test set F1 score**
Optimized architecture	0.848	0.842
U-Net original configuration [63]	0.7402	0.7557

Table 5.3: Comparison of validation and test mean F1 scores between the optimized network architecture and the original U-Net architecture [63].

Two key observations can be drawn from the architectural hyperparameter study:

1. **Number of convolutions per layer:** The most influential parameter is the number of convolutional layers per block. Even though the best-performing configuration features a network depth of five, reducing the depth to four while maintaining the same convolutional pattern leads to only a slight degradation in performance. This suggests that simply increasing network depth does not necessarily yield significant performance gains once a sufficient capacity is achieved.

2. **Initial number of filters:** The initial number of convolutional filters has only a minor effect on the final performance compared to the number of convolutions per layer.

Effect of kernel size. The kernel size was kept constant during the hyperparameter optimisation for computational efficiency reasons. To assess its impact, the best configuration obtained with a kernel size of three was compared to a model using a kernel size of five. As illustrated in Figure A.3, while a marginal improvement of less than 0.04% in the F1 score was observed when using a kernel size of five, it came at the cost of approximately 1.5 times greater computational time. Consequently, a kernel size of three was selected for the final model configuration.

Final hyperparameter configuration. The final set of hyperparameters used for subsequent experiments is summarised in Table 5.4.

	Depth	**Convolutions per layer**	**Initial filters**	**Kernel size**	**Learning rate**	**α**	**γ**	**Batch size**
Optimized hyperparameters	5	[3,4,5,5,5]	32	3	1.15×10^{-4}	0.8	2.0	32

Table 5.4: Final set of hyperparameters selected after the two-stage optimisation process.

5.1.5 Performance on the test set

The final hyperparameter configuration, obtained through the optimisation process, was used to assess the model's generalisation capabilities. This evaluation focuses specifically on the test set, which consists of previously unseen samples drawn from a separate region of the parameter space. As such, it serves as a rigorous benchmark for evaluating the model's ability to extrapolate beyond the training distribution.

Figure 5.4 illustrates the distribution of the F1 score over the test set. The results show a relatively uniform F1 score across different samples, indicating that the model does not exhibit any significant degradation in prediction accuracy within specific regions of the test domain. This homogeneity in performance suggests that the model has successfully learned generalisable features, enabling robust reconstruction of unstable branches across varying dynamical regimes. The average F1 score over the test set is 84.2 %.

The consistent performance on the test set implies that the training dataset provided sufficient coverage of the system's nonlinear behaviour, allowing the network to generalise to unseen configurations. Although the test samples correspond to systems with distinct dynamics, the model exhibits strong predictive capability.

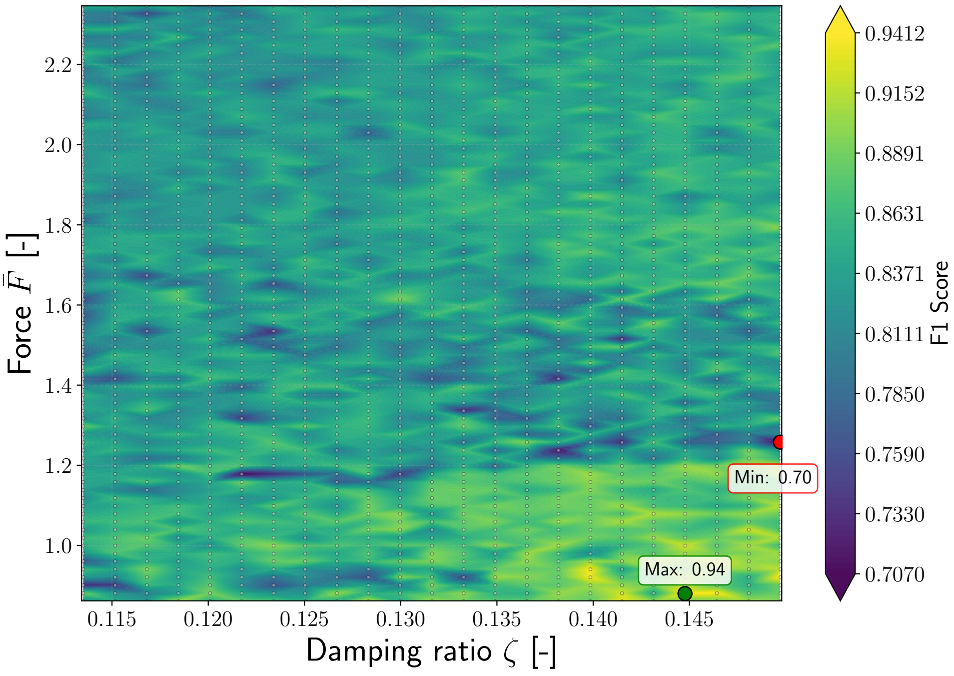
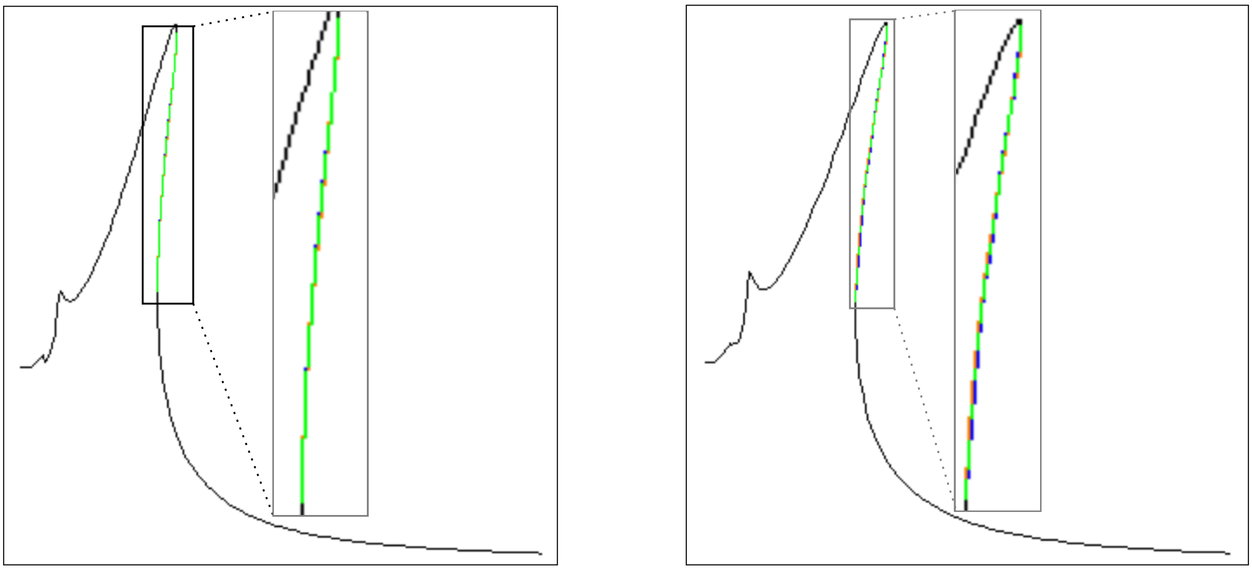


Figure 5.4: Distribution of the F1 score over the test set. Minimum and maximum values are highlighted.

Further insight into the model's behaviour is provided in Figure 5.5, which shows two representative examples from the test set: the samples with the highest and lowest F1 scores. In both Figures, the input data is shown in black. Correctly predicted pixels (true positives) are marked in green, missed pixels (false negatives) in blue, and incorrectly added pixels (false positives) in orange.

Even in the worst-performing case, the misclassified pixels are spatially adjacent to the ground truth regions. This proximity suggests that the model, despite local inaccuracies, retains a strong grasp of the qualitative structure of the unstable branch. The predictions remain relevant and interpretable, affirming the model's robustness in identifying bifurcated regions, even under challenging scenarios.



(a) Best prediction: $\zeta = 14.5\%$, $\bar{F} = 0.88$ **(b)** Worst prediction: $\zeta = 15\%$, $\bar{F} = 1.26$

Figure 5.5: Best (a) and worst (b) predictions on the test set in terms of F1 score. Black: input data. Green: true positives. Blue: false negatives. Orange: false positives.

To reconstruct a continuous representation of the predicted unstable branch from the discretised CNN output, an interpolation procedure was devised based on the output probability grid. Figures 5.6 and 5.7 illustrate this process for the worst- and best-case predictions, respectively.

Due to the use of the Bresenham line algorithm for discretising continuous frequency response curves into pixel grids, multiple pixels can be activated[1] along the amplitude axis for a single frequency bin. This issue is particularly pronounced near fold bifurcations, where the local slope of the frequency response is high. Given an image resolution of 320 × 320 over a fixed bandwidth, the frequency resolution is the bandwidth over the image resolution. In regions of steep slope, this resolution results in a spread of the curve across multiple amplitude bins for a single discrete frequency value.

To mitigate this, a weighted interpolation approach is employed: for each frequency bin, the corresponding amplitude bins with nonzero activation are averaged using their predicted probabilities as weights. This process yields one interpolated amplitude value per frequency, effectively converting the sparse pixel data into a smooth curve. However, when multiple amplitude bins are activated for a single frequency, the model often assigns the highest probability to the central pixel, regardless of the true curve position.

Despite these limitations, the quantitative evaluation shows that the maximum relative amplitude error between the interpolated and true unstable branches is only 4.5% in the worst case. Even under these worst-case conditions (Figure 5.6), the interpolated unstable branch remains qualitatively close to the ground truth obtained via HB continuation. In the best-case scenario (Figure 5.7), the predicted curve closely matches the ground truth, with a maximum relative amplitude error of only 2.48%.

---

[1] Having a probability greater than the probability decision threshold.

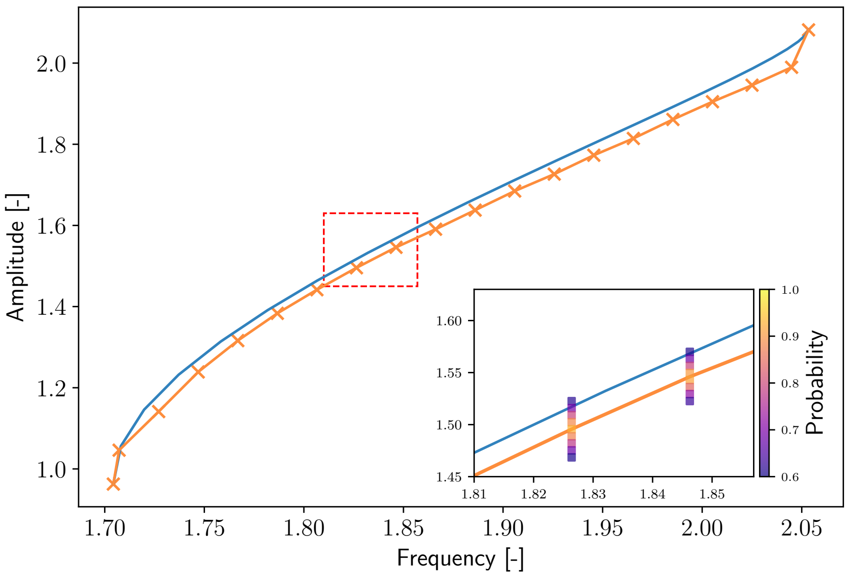


Figure 5.6: Interpolation of the predicted unstable branch (orange) in the worst-case scenario, based on the output probability grid. Ground truth from HB continuation is shown in blue.

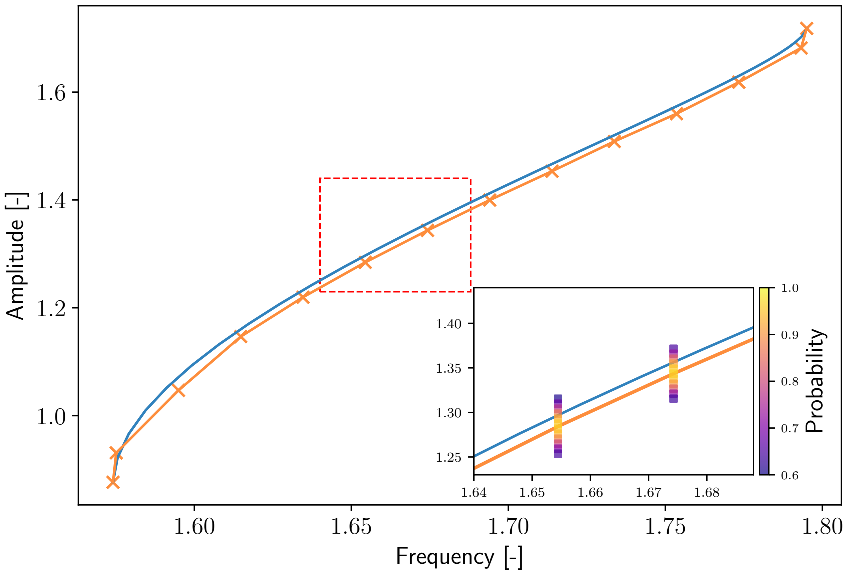


Figure 5.7: Interpolation of the predicted unstable branch (orange) in the best-case scenario. The HB continuation ground truth is shown in blue.

Two possible improvements can be proposed to refine the prediction and interpolation processes:

- **Higher resolution:** Increasing the grid resolution would reduce discretisation artefacts and minimise vertical spreading in regions of steep slopes. This would lead to more accurate pixel-level representations of the frequency response curve, thus improving the final interpolation.
- **Improved data encoding:** Rather than using Bresenham lines, which assign uniform weight to all pixels, one could adopt anti-aliased line drawing (e.g., Xiaolin Wu's algorithm), which computes pixel intensities proportional to the distance from the ideal line and yields smoother, sub-pixel–accurate representations without large aliasing artefacts.

While increasing the resolution is straightforward to implement, the adoption of anti-aliased encoding would require a more fundamental change to the pipeline. In particular, the current binary classification framework and loss function would no longer be applicable and would need to be redefined to accommodate continuous-valued targets. Nevertheless, this remains a promising direction for future investigation, as it could potentially improve predictive performance.

5.2 Prediction of secondary resonances

This section investigates the prediction of both *superharmonic* (3:1 and 2:1) and *subharmonic* (1:3 and 1:2) resonances in the Duffing oscillator.

5.2.1 Prediction of superharmonics

The prediction of the superharmonics of the Duffing oscillator follows the same general approach as presented in the previous chapter on fundamental resonance prediction. The image encoding strategy for the superharmonics was discussed in Section 4.4 and essentially involves encoding each superharmonics individually with an experimentally reproducible window. The parameter grid used for dataset generation is identical to that of the fundamental case. However, it is important to note that not all parameter configurations exhibit superharmonic responses, and their presence depends sensitively on the system's nonlinear characteristics.

The complete tracking of the 2:1 superharmonic branch in the Duffing oscillator typically requires the use of advanced techniques such as branch switching since standard continuation methods follow the primary (main) solution branch and fail to switch on another branch at a bifurcation point [66]. Without such techniques, the 2:1 superharmonic response may remain entirely undetected in numerical continuation. However, implementing a robust and general-purpose branch switching algorithm is non-trivial, particularly in the context of dataset generation, where consistency and automation across a wide range of parameter configurations are essential.

To circumvent this limitation, a small quadratic stiffness term was introduced into the system, effectively removing the isolated unstable segment typically found below the 2:1 superharmonic. As a result, the continuation algorithm can naturally follow the 2:1 branch without requiring branch switching techniques. The added quadratic nonlinearity is kept sufficiently small so as not to qualitatively affect the overall system dynamics. This modification was selected for its simplicity and practical robustness, particularly in the context of automated dataset generation. Figure 5.8 illustrates the dataset constructed for the superharmonic prediction task.

The two samples from the dataset that have the highest and lowest F1 scores are shown in Figure 5.9. This figure presents the true positive, false negative, and false positive pixel counts for each sample. In both cases, the scenarios involve 3:1 superharmonic resonances. The sample with the highest F1 score achieved an impressive 100%, while the sample with the lowest score recorded 47.6%.

As previously observed for the prediction of the fundamental resonance, similar conclusions can be drawn regarding the superharmonic predictions. Even in the most unfavourable cases, i.e., those associated with the lowest F1 scores, the misclassified pixels tend to be adjacent to the correctly predicted regions. Once the output is re-interpolated, this spatial proximity results in minimal deviation from the ground truth. This suggests that the model, when applied across the test set, maintains a consistent level of predictive accuracy after interpolation.

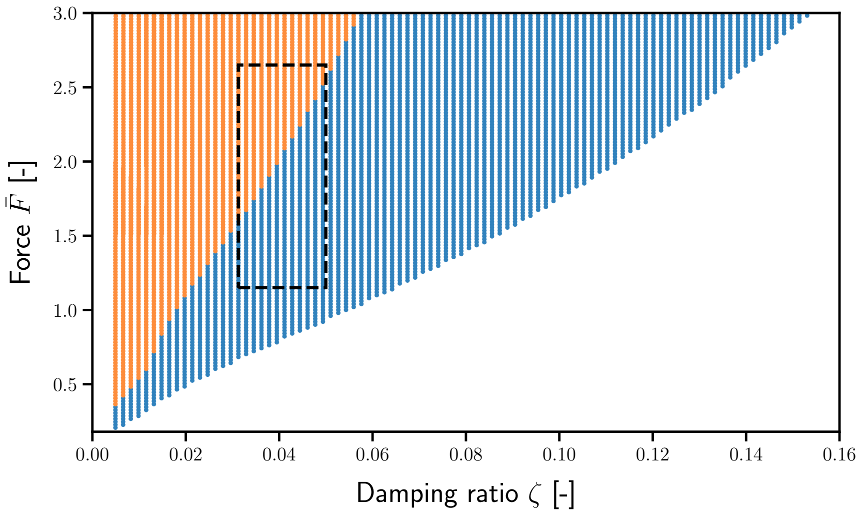
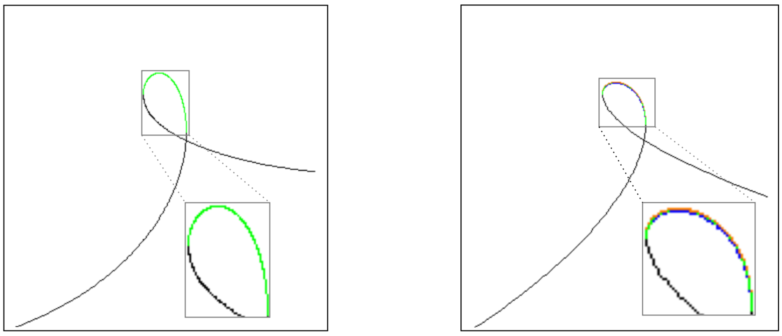


Figure 5.8: Visualisation of the Duffing subharmonic dataset. Blue points indicate parameter configurations yielding only a 3:1 subharmonic sample. Orange points indicate configurations yielding both 3:1 and 2:1 subharmonic samples. The black dashed rectangle marks the test set region.

Moreover, these observations highlight an important limitation of standard binary classification metrics, such as the F1 score, in the context of this problem. While binary classification metrics such as the F1 score provide a pixel-wise assessment of accuracy, they do not account for the smoothness or spatial continuity of the predicted regions. As a result, these metrics may not reflect how closely the interpolated prediction aligns with the true frequency response curve. A more meaningful assessment of the model's practical performance may thus be obtained by directly interpolating the predictions and comparing them to the ground truth obtained via HB continuation, using a global metric such as the relative error. This would provide a more representative measure of the model's capability to reconstruct the frequency response curve with fidelity.



(a) $\zeta = 10.4$ %; $\bar{F} = 2.18$

(b) $\zeta = 9.8$ %; $\bar{F} = 1.5$

Figure 5.9: Best (a) and worst (b) predictions on the test set in terms of F1 score. Black: input data. Green: true positives. Blue: false negatives. Orange: false positives.

5.2.2 Prediction of subharmonic isolas

In Section 2.1.3, the FRCs of the Duffing oscillator highlighted the occurrence of subharmonic resonances which appear as isolated solution branches, commonly referred to as *isolas*, detached from the main resonance path. As shown in Figure 2.4, increasing the forcing amplitude can lead to the sudden emergence of such isolas. Predicting their onset is particularly challenging due to the sensitivity of their appearance to system parameters and initial conditions. This behaviour is closely linked to the structure of the system's basins of attraction, which dictate convergence toward different coexisting periodic solutions. A proper understanding of the emergence of these isolas is therefore essential for fully characterising the nonlinear dynamics.

The dataset partition follows the same strategy used for predicting the fundamental resonance: 80%

of the data is used for training, 10% for validation, and 10% for testing. However, the problem formulation differs slightly. Here, the objective is to assess whether a model can accurately predict the existence of subharmonic isolas in previously unseen systems. To enforce this, the test set is constructed by selecting a continuous range of damping ratios across the entire span of forcing amplitudes. Consequently, the test data represents entirely new systems not present in the training set. This partitioning is visualised in Figure 5.10.

To ensure a sufficient number of samples containing isolas, the frequency shift threshold was increased to 1000%. This choice is motivated by the fact that subharmonic branches generally appear at higher frequency shifts. The frequency range is thus extended from 0 up to 10 times the linear resonance frequency.

As illustrated in Figure 5.10, the dataset includes three types of samples: those without subharmonic isolas, those with a 1:3 subharmonic, and those featuring both 1:3 and 1:2 subharmonic resonances. The goal is to evaluate the model's ability not only to detect the presence of isolas but also to accurately distinguish between these scenarios. The model must also be capable of reliably predicting the absence of secondary resonances.

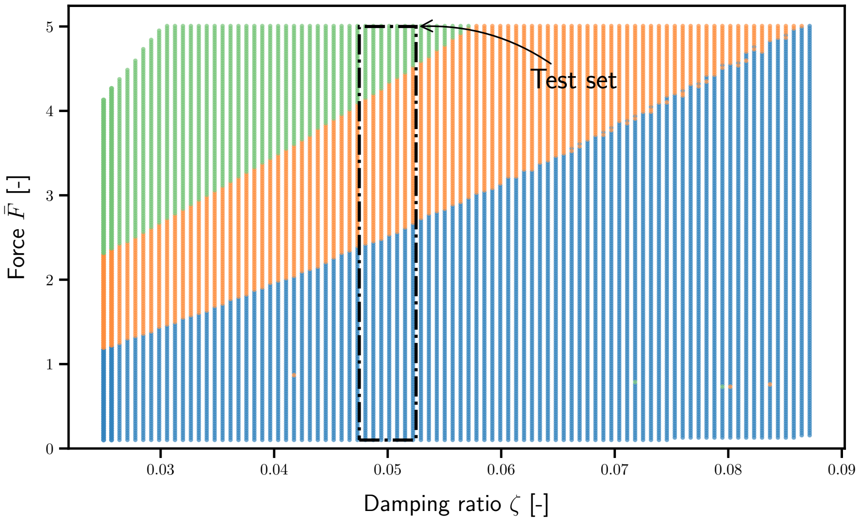


Figure 5.10: Representation of the dataset for Duffing subharmonics. Blue dots indicate no subharmonics, orange corresponds to the presence of the 1:3 subharmonic, and green to both 1:2 and 1:3 subharmonics.

The model architecture and training strategy are identical to those used for the fundamental resonance prediction. However, The metrics used for evaluating the fundamental resonance prediction are no longer appropriate in this context. To better assess the model's performance, the false discovery rate (FDR) and the false negative rate (FNR) are used instead and are defined as follows:

$$\mathrm{FDR} = \frac{\mathrm{FP}}{\mathrm{FP}+\mathrm{TP}}, \quad \mathrm{FNR} = \frac{\mathrm{FN}}{\mathrm{FN}+\mathrm{TP}}. \tag{5.1}$$

The FDR measures how often the model incorrectly predicts an isola when none exists, while the FNR quantifies the frequency with which the model fails to detect isolas that are actually present. These two metrics thus provide complementary insights into the model's performance in predicting secondary resonances. Figure 5.11 depicts the FDR and the FNR on the test set with the model hyperparameters 5.4.

High FDR or FNR values near the transition boundaries, such as between cases with no isolas and those featuring a 1:3 isola, or between regions with only a 1:3 isola and those with both 1:2 and 1:3 subharmonic isolas would indicate that the model either fails to detect isolas when they are present or incorrectly predicts them when they are absent. However, the results show that this is not the case: both FDR and FNR remain well below unity across the test set. These consistently low values confirm the model's ability to accurately capture the onset of isolas and to reliably differentiate between the various subharmonic resonance regimes. This demonstrates that the model is effective at predicting the parameter regions where subharmonic isolas emerge.

This predictive capability will be further assessed in the next chapter using experimental data. Nevertheless, based on the numerical results obtained so far limited to the SDOF Duffing oscillator, the model appears to generalise well and successfully identify the presence of subharmonic isolas across different system configurations.

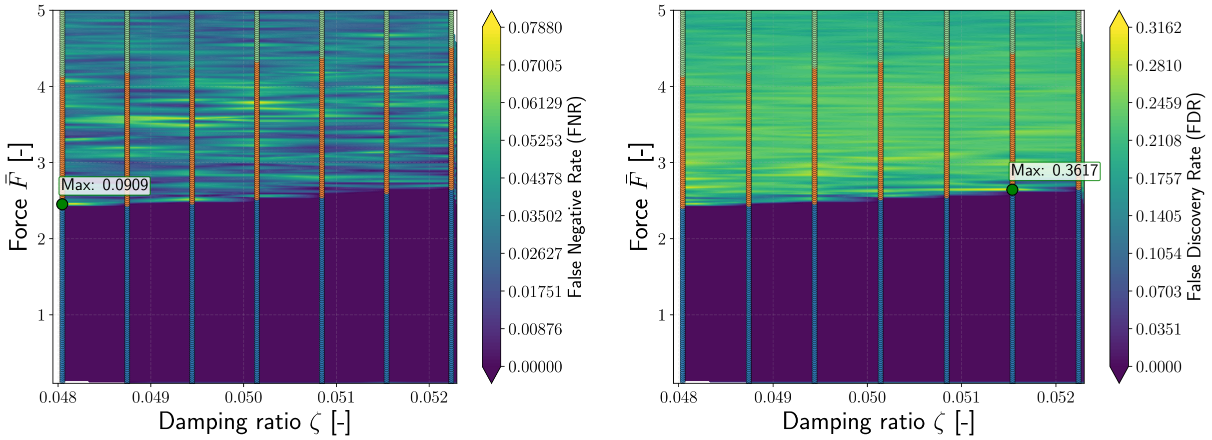
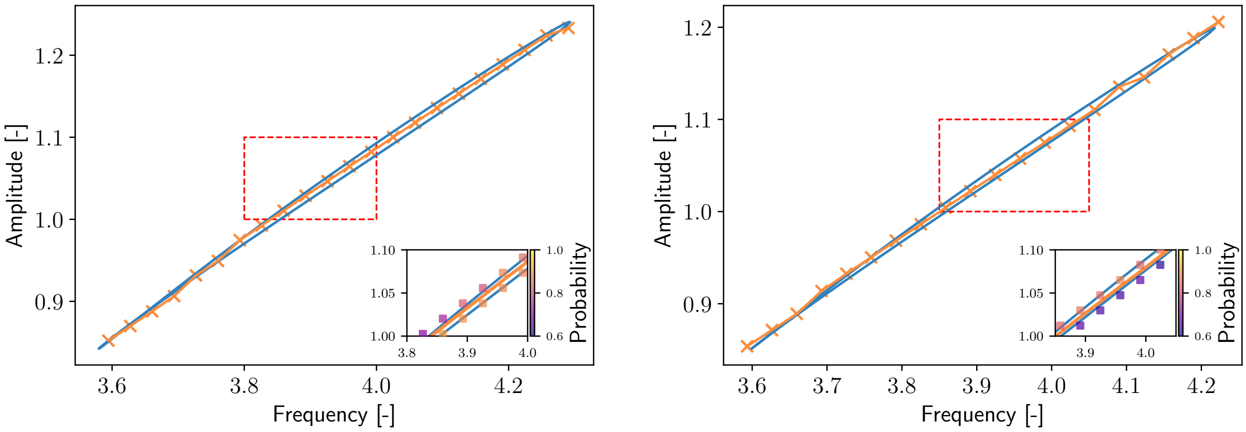


Figure 5.11: FNR in (a) and FDR in (b) are shown for the test set. Background colours reveal the spatial distribution of FNR and FDR values across the parameter space (ζ,$\bar{F}$). Color bars indicate scales for each metric. Blue dots represent samples without isolas, orange dots indicate those with a 1:3 subharmonic, and green dots show samples with both 1:2 and 1:3 subharmonics.

The best and worst predictions in terms of F1 score are presented in Figure 5.12, with corresponding F1 values of 0.98 and 0.77, respectively. These figures compare the ground truth obtained from HB continuation with the interpolated probability maps, following the same methodology used for the fundamental resonance case. Notably, in both scenarios, the subharmonic isola is very narrow, activating at most two pixels along the amplitude axis. As a result, when the probability interpolation is computed by averaging over frequency bins, the predicted isola tends to be centred between the stable and unstable branches.

It is important to emphasise that the predicted isola represents the complete solution, both stable and unstable branches. To further illustrate the structure of the predicted response, Figure 5.13 shows the contour lines delineating the predicted upper and lower boundaries of the isola.

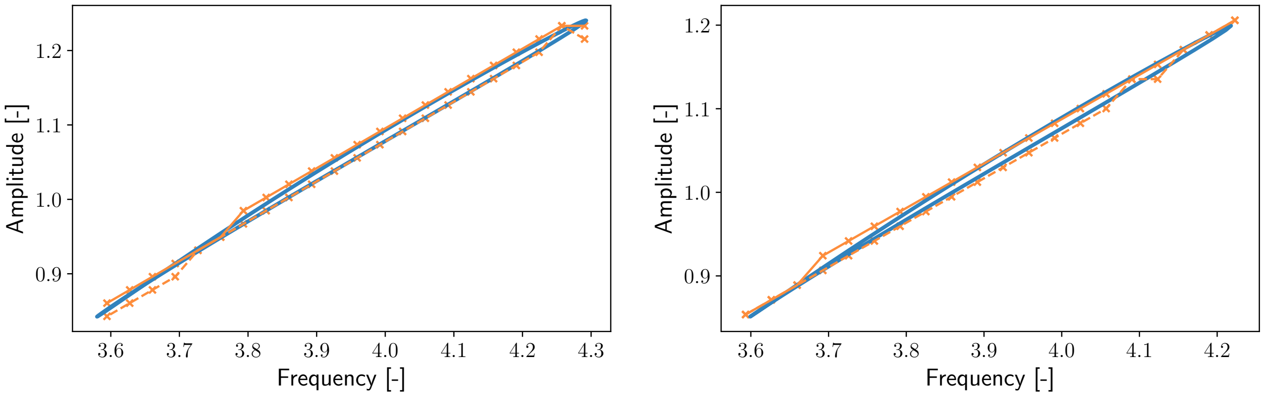
In both best- and worst-case scenarios, the frequency range over which the isola is predicted aligns accurately with the ground truth. Moreover, the consistency in the predicted contours of the upper and lower branches reflects the method's robustness. Nonetheless, as observed previously for the fundamental resonance prediction, improvements can be envisioned through enhanced image resolution and refined encoding schemes. Despite this, the current predictions provide a reliable and informative depiction of the nonlinear subharmonic response.



(a) Best prediction ($\zeta = 0.05\%$; $\bar{F} = 2.54$)

(b) Worst prediction ($\zeta = 0.048\%$; $\bar{F} = 2.43$)

Figure 5.12: Interpolation of the best (a) and worst (b) case scenarios in terms of F1 score on the test set. Ground truth from HB continuation is shown in blue.



(a) Best prediction ($\zeta = 0.05\%$; $\bar{F} = 2.54$)

(b) Worst prediction ($\zeta = 0.048\%$; $\bar{F} = 2.43$)

Figure 5.13: Contours of the predicted isola branches in the best (a) and worst (b) test cases based on F1 score. The upper branch is indicated by a continuous orange line, the lower branch by a dashed orange line, and the HB continuation is shown in blue.

5.3 Chapter conclusion and discussion

Chapter conclusion

This chapter began with the prediction of the fundamental resonance of the Duffing oscillator, including a description of the dataset generation process and the partitioning strategy used to create the training, validation, and test sets. The hyperparameter optimisation procedure described in Section 3 was then applied, leading to the identification of optimal hyperparameters, including those associated with the focal loss.

Subsequently, the model's performance on the test set was assessed using these hyperparameters, with both the best- and worst-case scenarios based on the F1 score. An interpolation strategy was also introduced to reconstruct continuous representations of the predicted unstable branches from the discretised output, and possible improvements to this approach were discussed.

A similar methodology was applied to the prediction of the first superharmonic resonances of the Duffing oscillator (2:1 and 3:1). Even in the worst-case scenario, interpolation of the probability grid enabled accurate predictions when compared to the ground truth obtained via HB continuation.

Finally, the prediction of subharmonic resonances (1:2 and 1:3), which appear as isolated solution branches (isolas), was investigated. The partitioning strategy differed in this case to ensure the test set included entirely unseen systems, thereby evaluating the model's ability to generalise to new configurations. The results were conclusive: the model successfully predicted the onset of subharmonic isolas and distinguished between different resonance regimes. Both the best- and worst-case scenarios were illustrated, and the boundaries of the predicted response were compared to the ground truth obtained via HB continuation.

This chapter confirmed that the proposed machine learning framework is capable of predicting unstable branches in nonlinear systems based solely on stable input data. In this context, the numerical approach is validated. However, the scope of this chapter is restricted to a single system and to purely numerical simulations. The primary objective here was to establish, through numerical experiments, the feasibility of this approach. In the next chapter, the models developed will be applied to an experimental electronic Duffing oscillator to assess whether this methodology can also be extended to real-world setups.

CHAPTER 6

Experimental validation

While previous assessments of the model's capabilities relied solely on numerical tests, this section focuses on experimental validation to determine the viability of practical implementation. The experimental characterisation of nonlinear systems remains a challenging and complex task in engineering research.

Various methodologies have been developed to identify nonlinear characteristics from experimental data, including shaker tables [15], hammer impacts [47, 55, 85], and electrodynamic vibration exciters [1, 6]. Electrodynamic shakers are particularly valuable as they apply controlled forces to the oscillator of interest through user-defined voltage signals. These devices are preferred for their ability to produce consistent, repeatable experiments with diverse signal inputs, including sine waves, random excitations, and transients.

Despite their widespread application, shakers present a notable limitation: shaker-structure interaction can distort the applied force, introducing non-fundamental harmonic content in force measurements [16, 56]. This phenomenon must be taken into account when analysing experimental results. It is precisely for this reason that an electronic Duffing oscillator was chosen for the experimental validation. This approach eliminates complications arising from shaker-structure interactions, allowing for a focus exclusively on validating the methodology without introducing mechanical coupling artefacts.

This section details the experimental application of the different models previously trained on the Duffing oscillator to validate whether the methodology described earlier can be successfully implemented under real-world conditions. The experimental setup employs an electronic Duffing oscillator, which is an analogue electronic circuit designed to replicate the behaviour of a Duffing oscillator.

6.1 Electronic Duffing

The Electronic Duffing Oscillator [60], illustrated in Figure 6.1, represents a precise electronic analogue of the mechanical Duffing system. This sophisticated circuit has been specifically engineered to exhibit minimal dissipation while maintaining exceptionally strong nonlinear characteristics.

Thanks to the analogue circuitry design, which provides test points and offers a clear analogy between electronic and mechanical systems, precise system identification is achievable. As demonstrated in [60], parameter identification shows only approximately 5% deviation between theoretical and experimental values.

However, the electronic implementation exhibits three primary imperfections affecting sine sweep experiments. Voltage offsets at the input and output of the integrated circuits compromise oscillator symmetry. Operational amplifiers with finite open-loop gain produce imperfect integration, introducing phase and amplitude errors. Furthermore, the analog multipliers introduce calculation errors, resulting in an imperfect cubic nonlinearity. These imperfections collectively manifest as minor discrepancies between theoretical predictions and experimental measurements during sine sweep tests.

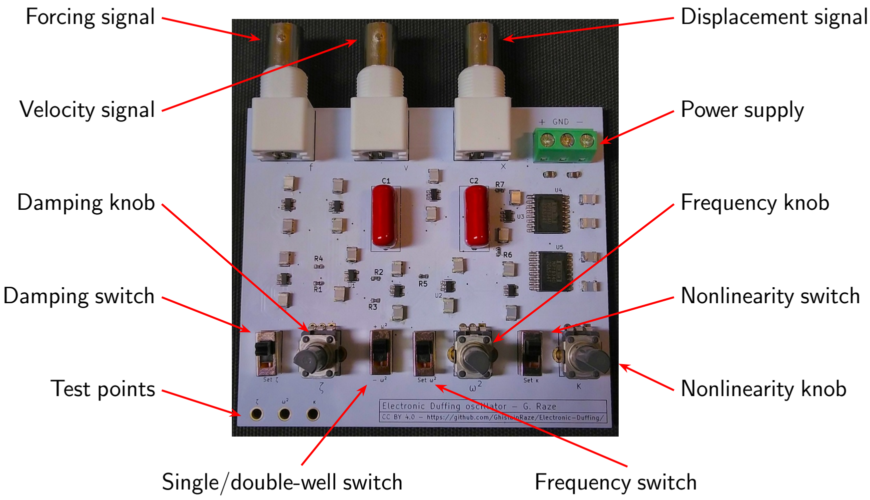


Figure 6.1: Setup of the electronic Duffing [60]. The input voltage is the force, and the output voltages correspond to the displacement and velocity of the oscillator.

6.2 Experimental measurement

The models trained to predict both the fundamental and secondary resonances have been applied to the electronic Duffing oscillator. Experimental data were collected via swept-up and swept-down sine tests over the same frequency ranges used during model training. Table 6.1 summarises the oscillator parameters.

Parameter Unit	m [s^2]	k_1 [–]	k_3 [V^{-2}]	c [s]	F [V]	$\bar{F}$ [-]	ζ [-]
Value	1.00×10^{-4}	1.68	0.983	4.9534×10^{-4}	2.00	0.91	1.91 %

Table 6.1: Parameters of the electronic Duffing oscillator.

To generate the model input, the envelope of the oscillator's response from swept-sine experiments is extracted. This envelope corresponds to the stable branches of the FRC. However, in the vicinity of bifurcations, the system experiences significant transient dynamics. These transients distort the true steady-state behaviour and, if retained, degrade the quality of the extracted envelope.

In particular, near superharmonic resonances, transient effects can manifest as pronounced overshoots in the measured response. These deviations from the steady-state amplitude can obscure the identification of the stable solution branches. To mitigate this issue, the data are carefully truncated near bifurcation points in order to isolate the true steady-state portion of the response envelope. Prior to truncation, an approximate estimation of the bifurcation frequencies is required to identify and exclude the transient regions with sufficient precision.

Two complementary strategies can be employed to enhance the accuracy of bifurcation localisation: reducing the sweep rate in swept-sine tests or utilising a stepped-sine approach. Decreasing the sweep rate reduces overshoots, yielding a closer approximation to the actual fold location. On the other hand, the stepped-sine method involves applying discrete frequency steps and allowing the system to reach steady-state at each step. This technique minimises transient effects and can achieve high-fidelity fold detection, particularly when applied over narrow frequency intervals with fine resolution.

However, while stepped-sine testing provides superior accuracy in localising bifurcations, it becomes increasingly impractical when extended over wide frequency ranges due to the substantial time required for system stabilisation at each frequency increment. In contrast, the swept-sine approach offers a more efficient means of characterising the full frequency range, particularly when the sweep rate is appropriately adapted to the resonance type and dynamic sensitivity.

For the fundamental and subharmonic isolas, the excitation frequency was swept over the ranges $[0,\ 6\omega_0]$ and $[0,\ 10\omega_0]$ respectively, using a sweep rate of 1 Hz/s to balance resolution and experiment duration. In contrast, for the more sensitive superharmonic responses (e.g., 3:1 and 2:1), a significantly slower sweep rate of 0.025 Hz/s was employed. Additionally, in the vicinity of bifurcations, further low-sweep-rate tests were conducted to enhance the fidelity of fold localisation.

Figure 6.2 shows the swept-up (blue) and swept-down (orange) responses around the fundamental resonance at $\tilde{F} = 2\,\mathrm{V}$. The black envelope corresponds to the portion of the response retained after trimming at the fold bifurcations. Similarly, Figure 6.3 presents the response envelopes near the 3:1 and 2:1 superharmonic resonances, depicted in Figures 6.3a and 6.3b, respectively. As before, the blue and orange curves show the raw swept-up and down responses, while the black curves indicate the trimmed envelopes used as input to the model.

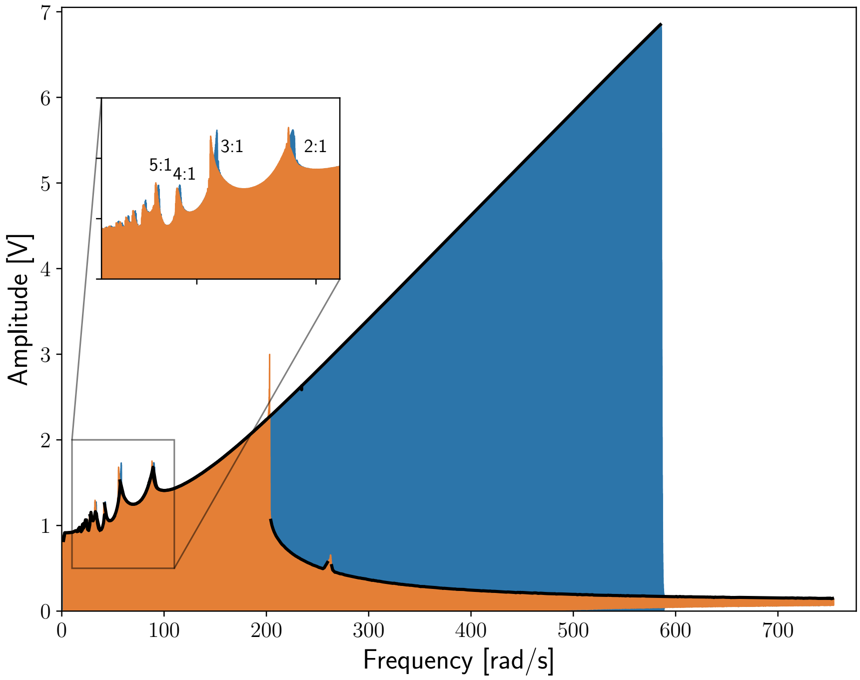
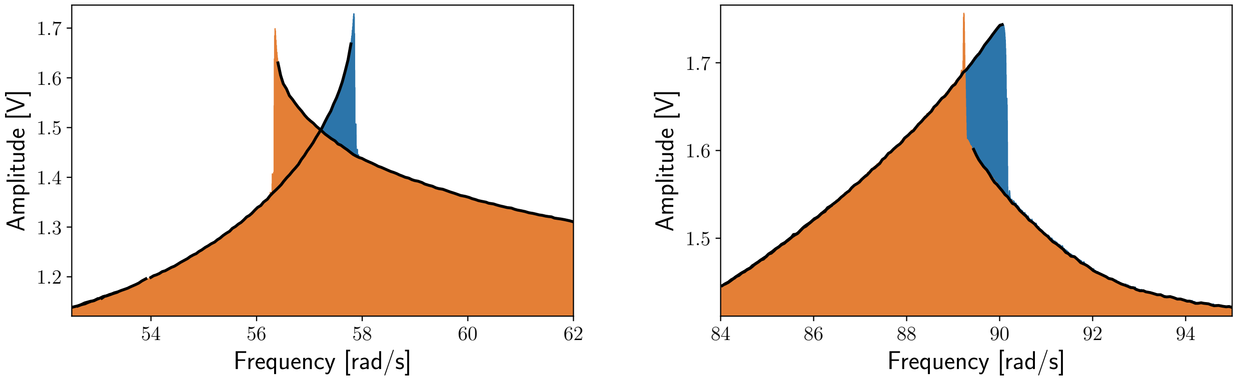


Figure 6.2: Swept-up (blue) and down (orange) sine tests responses of the electronic Duffing oscillator at $\tilde{F} = 2\,\mathrm{V}$. The envelope (black) is trimmed at fold bifurcations.

The experimental setup exhibits a response offset of approximately -0.115V. These unavoidable offsets break the system's inherent symmetry. As a result, resonances like the 2:1 superharmonic can emerge directly without symmetry-breaking bifurcations. The asymmetry, characterised by the symmetry factor $\bar{j}$, also modifies the bifurcation structure. As shown in [3], depending on the level of asymmetry, the 2:1 subharmonic branch can appear attached to the main resonance, as observed in the present case.



(a) 3:1 superharmonic

(b) 2:1 superharmonic

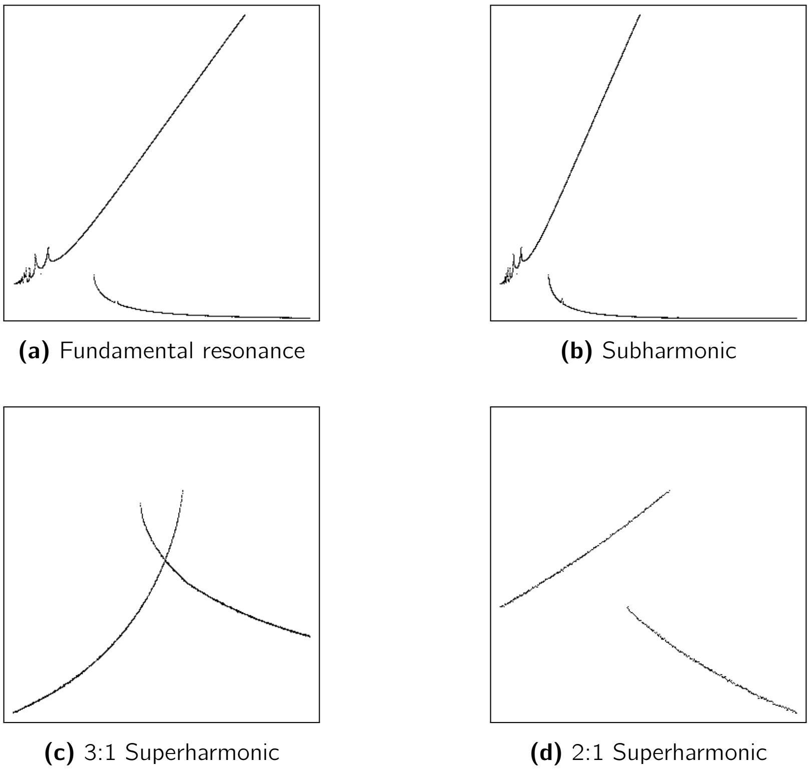
Figure 6.3: Swept-up (blue) and down (orange) sine tests responses around the (a) 3:1 and (b) 2:1 superharmonic resonances at $\tilde{F} = 2\,\mathrm{V}$. Black curves are the trimmed envelopes.

Furthermore, the superharmonic resonances appear at frequencies larger than respectively one-third and one-half of the natural frequency, highlighting the presence of extremely high nonlinearities in the system combined with high forcing.

To validate the predicted resonances, the ACBC [1] method is employed. This derivative-free, control-based technique facilitates the identification of both stable and unstable branches of FRCs, eliminating the need for numerical differentiation and thereby enhancing robustness to measurement noise. It achieves this by stabilising periodic orbits through feedback control while maintaining non-invasiveness via adaptive filtering, which ensures monoharmonic excitation.

Nevertheless, the standard ACBC framework may be insufficient for accurately capturing subharmonic and superharmonic resonances. To address this limitation, an extended formulation, referred to as extended-ACBC (x-ACBC) [71], is employed. This enhanced approach introduces a secondary continuation parameter, typically linked to a non-fundamental harmonic component of the response, allowing the procedure to target and even continue isolated solution branches.

Finally, the binary input images employed for model inference are shown in Figure 6.5. These input images were generated using only the envelopes from swept-sine responses and subsequently applying Bresenham's line algorithm. The encoding procedure strictly adheres to the same frequency and amplitude windows defined in Chapter 4, thereby ensuring consistency with the dataset generation methodology. Each image serves as input to a specific model dedicated to predicting a particular type of resonance. To ensure alignment with the training data, which were generated without any offset and therefore exhibit a decay to zero at high frequencies, the experimental data were corrected by subtracting the response offset. This effectively shifts the data and restores consistency with the model's input domain.



(a) Fundamental resonance **(b)** Subharmonic

(c) 3:1 Superharmonic **(d)** 2:1 Superharmonic

Figure 6.5: Binary images were generated using the Bresenham line algorithm applied to the envelope data from swept-sine experiments. These serve as inputs to predict: (a) fundamental, (b) subharmonic, and (c)–(d) superharmonic resonances.

6.3 Prediction of the fundamental resonance

The prediction of the fundamental resonance of the electronic Duffing oscillator for the parameters listed in Table 6.1 is carried out using the model trained on the fundamental dataset described in Section 5.1.2. The corresponding binary input image is shown in Figure 6.4a. In this experimental evaluation, the model used for inference was trained on the complete dataset, i.e. without a test/train partitioning, allowing the network to exploit all available data to enhance predictive performance.

As a reminder, the model outputs a probability map where each pixel indicates the likelihood of belonging to the unstable branch of the FRC. While a fixed decision threshold ($\tau = 0.6$) can binarise this map, a more robust alternative previously applied in Chapter 5.1.5 is to apply interpolation, which resolves ambiguities such as multiple predicted pixels within the same frequency bin.

Figure 6.6 shows the interpolated prediction (blue) alongside the experimental measurement obtained via x-ACBC (orange). The predicted curve aligns closely with the measurement across the entire frequency range. The largest relative amplitude error, still as low as 1.53%, occurs near the lower fold at $\omega \approx 205$ rad/s.

Additionally, the jump down and the calculated fold point with x-ACBC are closely related. This is important because the model was trained using numerical samples obtained through HB continuation, which means the unstable branch was separated at the actual bifurcation points. Although the jump down is near, it may not be precisely at the exact bifurcation point. However, with a sufficiently low sweep rate, the jump down remains close to the bifurcation. This high level of accuracy aligns well with the model's performance during numerical validation.

While this result is promising, it also suggests room for further refinement. As discussed in Chap-

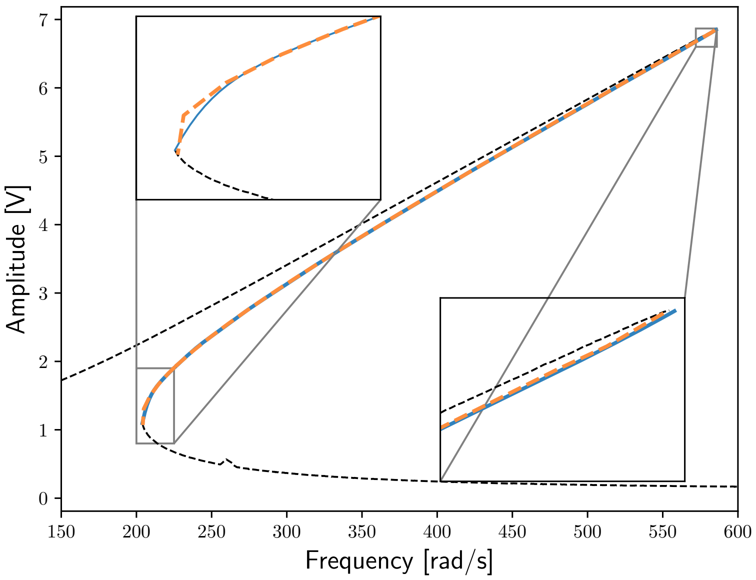
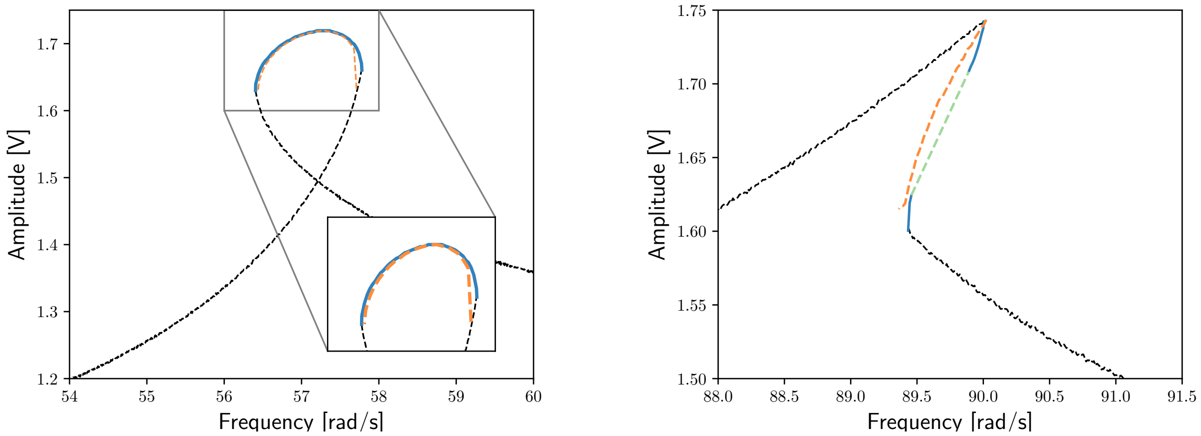


Figure 6.6: Prediction (—) and measurement via x-ACBC (—) of the FRC of the electronic Duffing oscillator for $F = 2$ V. The black dashed lines represent the envelopes extracted from the swept-sine tests.

ter 5.1.5, two key improvements were identified to enhance predictive performance: increasing the resolution of the input images and refining the encoding methodology.

6.3.1 Prediction of the superharmonics

Finally, the model trained on the superharmonic responses at ratios 3:1 and 2:1 is employed for inference using the input images shown in Figures 6.4c and 6.4d. The predicted responses, along with experimental measurements obtained via x-ACBC, are presented in Figures 6.7a and 6.7b.



(a) **(b)**

Figure 6.7: Predicted (—) and measured (—) superharmonic responses of the electronic Duffing oscillator for $F = 2$ V. The black dashed lines indicate the upper and lower envelopes of the response obtained from the swept-sine excitation tests, while the missed portion of the predicted 2:1 branch is shown in green.

The prediction of the 3:1 superharmonic shows excellent agreement with measurements obtained via x-ACBC, exhibiting a maximum relative amplitude error of less than 1%. Interestingly, the prediction appears even more accurate near the right fold bifurcation. This is due to the limited resolution of the x-ACBC measurements in that region, which did not fully capture the branch up to the fold point, effectively truncating it slightly beyond the bifurcation. In contrast, the prediction of the 2:1 superharmonic displays noticeable discrepancies with the measured response. In particular, the model fails to capture an entire branch of the response, which is highlighted by the dashed green line (representing the portion not predicted).

This limitation, however, can be explained by the nature of the 2:1 response. It emerges from a symmetry-breaking bifurcation induced by the presence of a static offset in the excitation. The static offset force F_{offset} can be estimated from experimental data by measuring the static displacement d_x in the absence of dynamic excitation. Assuming a restoring force with both linear and cubic components, the static equilibrium condition gives:

$$F_{\text{offset}} = k_1 d_x + k_3 d_x^3 \tag{6.1}$$

This provides a direct link between the observed static deformation and the equivalent static force acting on the system. Applying the change of variable $x(t) = y(t) + d_x$ to the original equation of motion:

$$m\ddot{x} + c\dot{x} + k_1 x + k_3 x^3 = F \sin(\omega t) \tag{6.2}$$

and simplifying leads to:

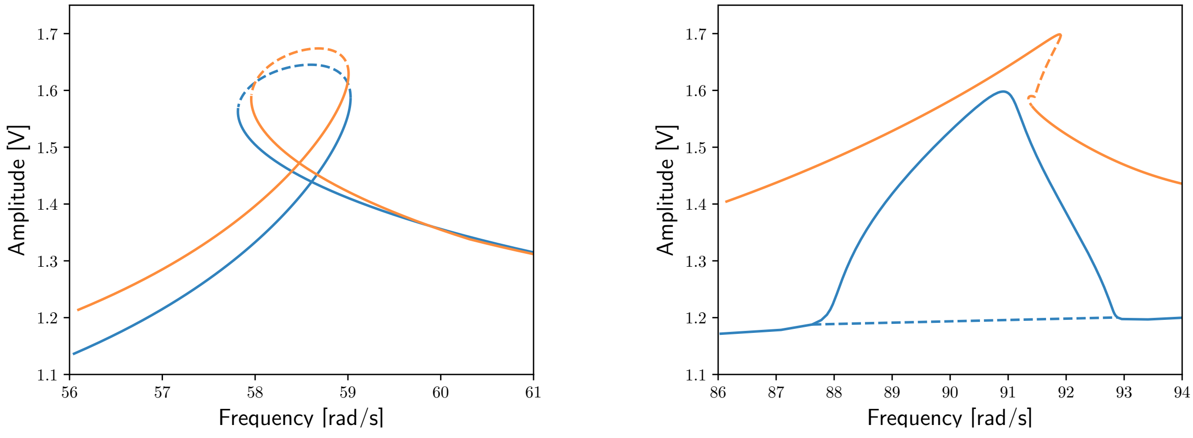
$$m\ddot{y} + c\dot{y} + \left(k_1 + 3k_3 d_x^2\right) y + 3k_3 d_x y^2 + k_3 y^3 = F \sin(\omega t) \tag{6.3}$$

This transformation reveals the influence of the static offset: the effective linear stiffness becomes $k_{\text{eff}} = k_1 + 3k_3 d_x^2$, slightly shifting the natural frequency, while the appearance of a quadratic nonlinear term $3k_3 d_x y^2$ introduces asymmetry into the system.

These modifications have a significant impact on the system's nonlinear dynamic behaviour. The discrepancy observed between the predicted and experimentally measured 2:1 superharmonic FRC can largely be attributed to the presence of this static offset. Figures 6.8a and 6.8b illustrate the FRC obtained via HB continuation for the superharmonics 3:1 and 2:1, both with (orange) and without (blue) the estimated offset.

In the absence of an offset, the 2:1 superharmonic does not contain a fold bifurcation. This further explains why the model failed to predict this feature: such configurations are simply not represented in the training database. Moreover, the shape of the 3:1 superharmonic also changes noticeably in the presence of the offset. Interestingly, the model successfully captured the correct shape of the 3:1 response despite not having been explicitly trained on data exhibiting this behaviour. This highlights the model's strong generalisation capability.

Other sources of uncertainty, such as nonlinear damping and deviations from a purely cubic restoring force due to tolerances in the electrical components, may also contribute to the mismatch. Nonetheless, the primary source of the observed differences remains the presence of the offset.

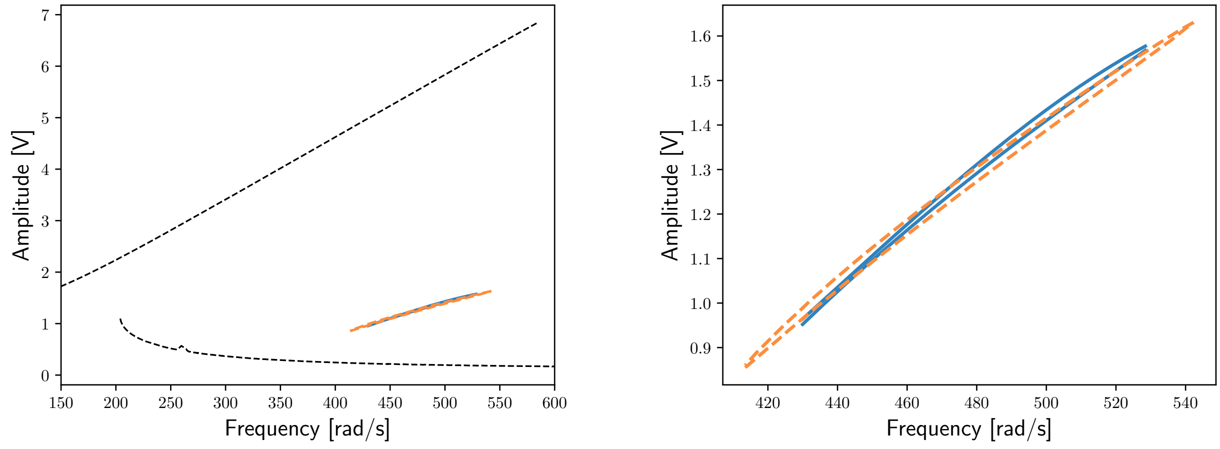


(a) **(b)**

Figure 6.8: FRCs around the 3:1 (a) and 2:1 (b) superharmonics of the equivalent electronic Duffing oscillator parameters ($\zeta = 1.91\%$; $\bar{F} = 0.91$) computed using HB continuation. The results are shown both with (—) and without (—) the estimated static offset in the restoring force. Solid lines indicate stable branches, while dashed lines represent unstable branches.

6.3.2 Prediction of the subharmonics

Finally, the prediction of the subharmonic resonances follows the same methodology as the one employed for the fundamental and superharmonic resonances. Inference is performed using the input image shown in Figure 6.4b. The predicted subharmonic responses are presented in Figure 6.9a, alongside the experimental measurements obtained via the x-ACBC method. For reference, the swept-sine measurements around the fundamental resonance are also displayed in dashed black lines. Figure 6.9b provides a detailed view focused on the 1:3 subharmonic resonance.



(a) **(b)**

Figure 6.9: Predicted (—) and measured (—) subharmonic responses of the electronic Duffing oscillator for $\tilde{F} = 2$ V. The black dashed lines correspond to the upper and lower envelopes extracted from the swept-sine excitation tests.

The predicted 3:1 subharmonic resonance closely matches the experimental measurements. The slight discrepancies observed can be attributed to the presence of a static offset and other system uncertainties, which cause deviations from the ideal Duffing behaviour on which the model was trained.

Nevertheless, the predictions remain strong and consistent with the trends discussed for the fundamental resonance. They also align well with the observations made in the numerical validation section.

6.4 Complete predicted FRC

To conclude, Figure 6.10 presents the FRC of the electronic Duffing oscillator for $\tilde{F} = 2$ V. This global FRC is assembled using only the swept-sine envelopes (black) and the predictions of the different models corresponding to the fundamental, subharmonic, and superharmonic components (blue).

The resulting figure provides a comprehensive visualisation of the system's nonlinear response landscape, combining experimentally accessible data with the model's predictions of the inaccessible unstable branches with standard EMA techniques. This illustrates the potential of the proposed methodology to enhance experimental system identification by enabling the reconstruction of complete FRCs.

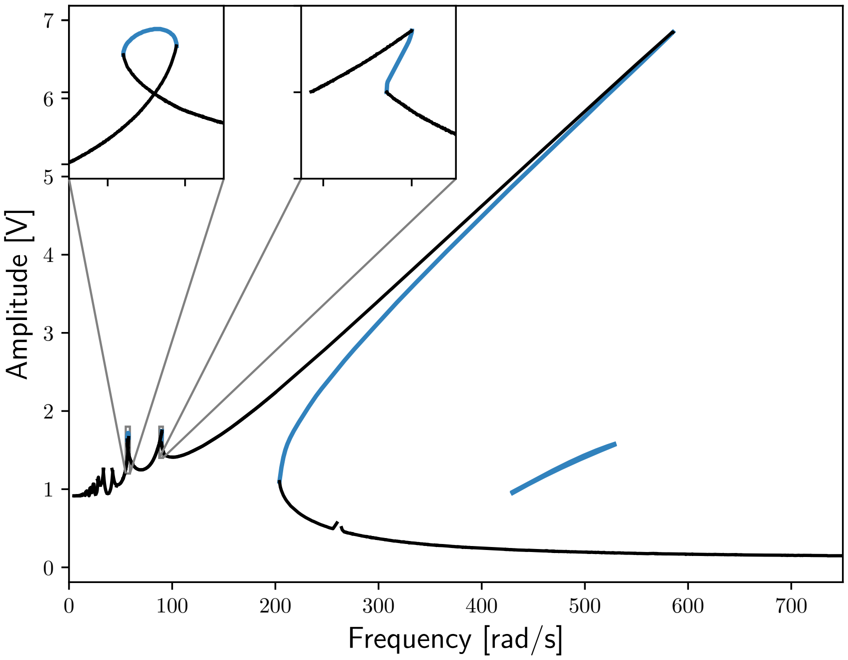


Figure 6.10: Reconstructed FRC of the electronic Duffing oscillator at $\tilde{F} = 2$ V. The blue curve represents the model predictions for fundamental, subharmonic, and superharmonic components, and the black curves correspond to the envelopes extracted from swept-sine measurements.

6.5 Chapter conclusion and discussion

Chapter conclusion

This chapter has demonstrated the practical applicability of the proposed ML methodology for predicting FRC using an experimental electronic Duffing oscillator as a testbed. The experiments were designed to validate the model's capability to generalise from purely numerical training data to real-world measurements.

The electronic Duffing oscillator offers a clean and controllable environment, thereby avoiding the common complications associated with mechanical experiments, such as shaker-structure interaction. Nonetheless, some imperfections inherent to the analogue implementation, such as offset voltages, non-ideal components, and analogue computation errors, introduce deviations from the ideal Duffing behaviour used during training.

Despite these challenges, the model showed remarkable robustness and generalisation ability. The prediction of the fundamental resonance was highly accurate, with only minor discrepancies near the fold points. Predictions of the 3:1 superharmonic and 1:3 subharmonic resonances also aligned well with experimental results.

A notable limitation arose in predicting the 2:1 superharmonic response. The model failed to capture a portion of this branch, which was shown, through numerical continuation with and without offset, to emerge only when symmetry is broken, in this case by a static offset. Since such offset-induced responses were not included in the training dataset, this behaviour is outside the learned domain of the model. Nevertheless, this finding suggests a potential extension of the dataset to include asymmetric cases.

In summary, the experimental validation confirms that the trained models are capable of effectively predicting both fundamental and secondary resonances, even in the presence of measurement imperfections. The results are consistent with the conclusions drawn from numerical validation and underscore the model's ability to generalise beyond idealised training conditions. The next chapter will extend the training data to a broader class of nonlinearities, aiming to further enhance the robustness and predictive power of the framework.

CHAPTER 7

Generalisation across multiple oscillators

The preceding chapters have focused exclusively on the Duffing oscillator, utilising both numerical simulations and experimental measurements to demonstrate the ability of deep learning models to predict its fundamental and secondary resonances accurately.

Although the Duffing oscillator offers a rich and well-studied example of nonlinear dynamics, it represents only a specific subclass within the broader category of nonlinear oscillators. To evaluate the generalisation capabilities of the models, this chapter extends the analysis to two additional systems: the Helmholtz-Duffing (H-D) oscillator and the piecewise linear (PWL) stiffness oscillator. As introduced in Chapter 2, these systems were selected for their distinct nonlinear characteristics and frequency response profiles. The H-D oscillator combines quadratic and cubic stiffness terms, resulting in a softening-to-hardening transition in its response, whereas the PWL oscillator is characterised by a non-smooth restoring force and, consequently, non-smooth dynamics.

The inclusion of these two systems enables the examination of the effect of training on a more diverse set of nonlinear systems. This chapter initially focuses on predicting the fundamental resonance, which serves as a consistent benchmark for assessing whether the incorporation of heterogeneous training data improves, degrades, or leaves the model's predictive accuracy unchanged. To this end, a new model is trained on a composite dataset that includes frequency response data from the Duffing, Helmholtz-Duffing, and piecewise linear oscillators. Its performance is then evaluated on the same numerical Duffing test set introduced in Chapter 5, enabling a direct comparison with a model trained solely on Duffing-type dynamics.

Subsequently, the model is tested on experimental data from the electronic Duffing oscillator described in Chapter 6 to investigate whether training on multiple nonlinear systems enhances robustness to real-world measurement imperfections. Finally, the model is evaluated on various nonlinear oscillators not present in the training set. This final assessment offers insight into the model's extrapolation capabilities and its ability to capture generalisable features of nonlinear dynamic systems.

7.1 Extension of the dataset

The dataset, originally composed exclusively of FRCs from the Duffing oscillator, has been expanded to include two additional nonlinear systems. As detailed in Section 2.3, the parameter spaces of these systems include the damping ratio ζ and the dimensionless forcing amplitude $\bar{F}$ parameters that are likewise defined for the Duffing oscillator. In addition, the H-D oscillator introduces the non-dimensional quadratic stiffness parameter β, while the PWL stiffness oscillator includes the stiffness ratio κ. Consequently, the parameter space becomes three-dimensional for both systems.

To ensure methodological consistency across all oscillator types, the window encoding strategy previously applied to the Duffing oscillator was retained. Similarly, the maximum allowable resonance frequency shift was fixed at 500 %. The ranges for the damping ratio ζ and the dimensionless forcing amplitude $\bar{F}$ were kept identical to those used in the Duffing dataset, thereby preserving uniform training conditions. The harmonic truncation order was also fixed at $N_H = 7$ for all oscillator types.

A particular challenge arises in the case of non-polynomial nonlinear restoring forces, such as those encountered in PWL systems. These forces are not infinitely differentiable and typically generate an infinite series of non-zero harmonics, which can lead to aliasing effects when discretised [39]. To mitigate such numerical artefacts, it is essential to choose a sufficiently large number of time samples N. An inadequate sampling rate can result in spurious oscillations. While increasing N improves spectral accuracy, it is at the expense of higher computational cost, which scales approximately as $N \log(N)$.

Table 7.1 summarises the parameter ranges employed in generating the datasets for the H-D and PWL oscillators. The design goal was to sample a broad and diverse set of system responses, thereby ensuring that a wide spectrum of FRC profiles is represented in the dataset.

Parameter	**Range of values**	**Number of points**
Damping ratio ζ [%]	$(0.5, 25)$	30
Dimensionless force $\bar{F}$	$(0.1, 3)$	30
	Helmholtz-Duffing	
Quadratic stiffness parameter β	$(0.5, 1.5)$	30
	Piecewise linear stiffness	
Stiffness ratio κ	$(2, 10)$	30

Table 7.1: Parameter values used for generating the Helmholtz-Duffing and PWL datasets.

To preserve balance across oscillator types, the dataset was constructed such that each system contributes one-third of the total samples. This results in a combined dataset of over 54,000 samples used for training.

7.2 Performance on the Duffing test set

To assess the influence of training data diversity on the model's predictive performance, a comparative evaluation is carried out using the same validation and test sets previously employed in Chapter 5. While the test and validation sets remain unchanged, the key modification in this analysis lies in the training data composition: the model is now trained on an extended dataset that includes not only Duffing oscillator samples but also samples from H-D and PWL stiffness systems.

This experimental setup enables a direct comparison between two models: one trained solely on Duffing oscillator data and another trained on an augmented dataset that includes multiple types of nonlinear oscillators. The objective is to determine whether incorporating a diverse set of nonlinear oscillators during training enhances, degrades, or preserves the model's ability to generalise to a specific oscillator type, in this case, the Duffing system.

Figure 7.1 presents the F1 score distribution obtained when evaluating the Duffing test set using the model trained on the augmented dataset. For reference, the performance of the model trained exclusively on Duffing samples is reported in Figure 5.4. The comparison highlights both differences in distribution and in the extrema of the F1 score across the test set.

Specifically, the minimum F1 score achieved by the model trained on the extended dataset is 62%, compared to 70% for the Duffing-only model. While this indicates a decrease in worst-case performance, the maximum F1 score remains comparable (95% versus 94%), and the corresponding samples for both extrema differ between models. Despite these local variations, the average F1 score across the test set remains effectively unchanged: 84.5% for the augmented model versus 84.2% for the Duffing-only model. This result suggests that the inclusion of additional nonlinearities during training does not compromise the model's ability to generalise to Duffing-type responses. Rather, it indicates that generalisation performance can be preserved, even in the presence of increased heterogeneity in the training data.

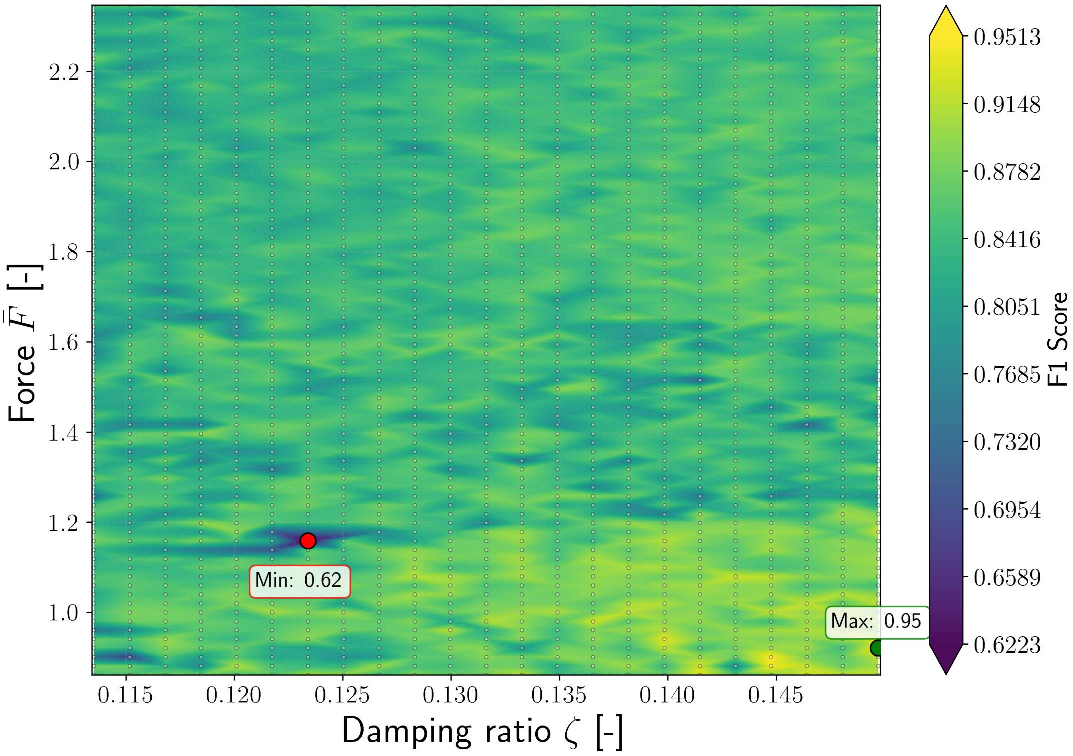


Figure 7.1: Distribution of the F1 score over the Duffing test set for the model trained on the augmented dataset. Minimum and maximum scores are annotated.

To further illustrate the model's behaviour, Figure 7.2 shows the prediction that resulted in the lowest F1 score on the test set. In this example, the misclassified pixels, both false positives and false negatives, are found to be spatially clustered. Consequently, even in this worst-case scenario, the local nature of the errors ensures that their overall impact is negligible once interpolation is applied.

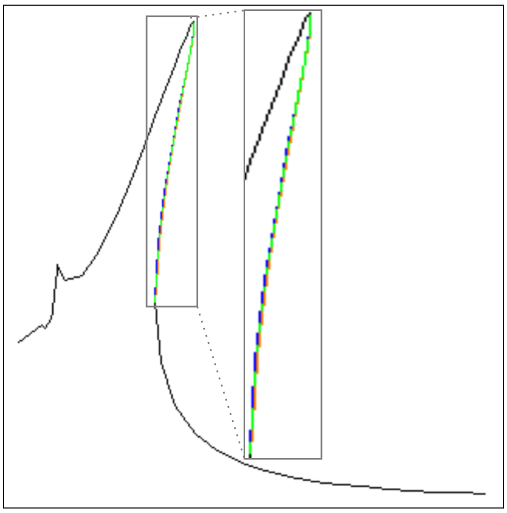


Figure 7.2: Worst prediction on the Duffing test set in terms of F1 score, for the model trained on the augmented dataset. Black: input data; green: true positives; blue: false negatives; orange: false positives.

7.3 Prediction on the electronic Duffing

In Chapter 6, the fundamental resonance and secondary resonances of the electronic Duffing oscillator were predicted using a model trained exclusively on Duffing oscillator data. Its performance was evaluated against experimental measurements obtained via the x-ACBC method. In the present analysis, the same prediction task, targeting the fundamental resonance, is revisited using a model trained on the complete augmented dataset. The aim is to examine whether training on a structurally diverse set of nonlinear oscillators affects the model's ability to predict the fundamental resonance of a real experimental setup. While the previous section addressed this question in a purely numerical context, the current investigation focuses on the experimental case.

To ensure a direct and controlled comparison, inference is conducted using the same input image as in Chapter 6 (Figure 6.4a), and the hyperparameters (see Table 5.4) remain unchanged. The resulting prediction of the unstable branch, along with the experimental measurement obtained via x-ACBC and the previously predicted unstable branch using the Duffing-only model, is shown in Figure 7.3.

As illustrated, the predicted curve deviates slightly from the one obtained using the Duffing-only model. The maximum relative amplitude error with respect to the x-ACBC measurement increased only marginally, from 1.53 % to 1.64 %. This maximum error is located near the low-amplitude fold at approximately $\omega \simeq 206.5$ rad/s. While the predicted unstable branch at lower frequencies lies slightly below the measured one, it rapidly aligns with the experimental measurement and matches it perfectly up to the resonance peak.

This systematic bias in the lower-frequency region could be attributed to the use of the Bresenham line algorithm. This algorithm, while efficient for line rasterisation, may introduce minor inaccuracies in regions with sharp curvature or low-amplitude features. A similar effect was observed during the numerical validation of the fundamental resonance in Chapter 5, indicating that this issue is recurrent across different use cases.

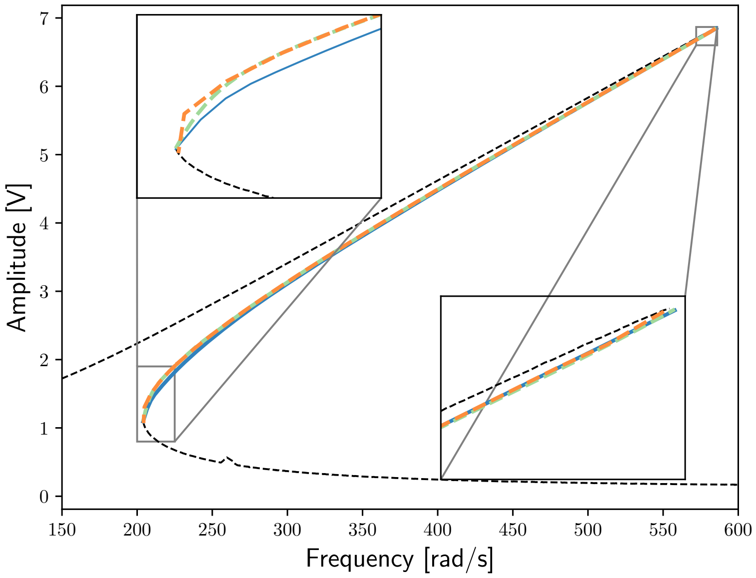


Figure 7.3: Prediction (—), measurement via x-ACBC (—), and prediction with Duffing-only model (—) of the FRC of the electronic Duffing oscillator for $F = 2$ V. The black dashed lines represent the envelopes extracted from swept-sine tests.

The consistency of the results obtained across both numerical and experimental evaluations demonstrates that training on a heterogeneous set of nonlinear oscillators does not diminish the model's ability to predict the behaviour of a specific system, such as the Duffing oscillator. Despite the inclusion of inherently distinct dynamics in the training data, predictive accuracy remains nearly unchanged.

7.4 Generalisation to unseen nonlinear oscillators

Up to this point, the analysis has focused exclusively on the Duffing oscillator, both through numerical simulations (Chapter 5) and experimental validation (Chapter 6). Even in the case of the electronic Duffing oscillator, discrepancies were observed between the measured data and the idealised Duffing model, attributed primarily to offset effects and component imperfections. Nevertheless, the dominant characteristics remained aligned with Duffing-type behaviour.

To evaluate the model's capacity to generalise beyond its training distribution, the following analysis considers nonlinear oscillators not included in the training set. The model trained on the augmented dataset is assessed based on its performance in predicting the fundamental resonance of a parametric family of systems governed by:

$$\bar{y}'' + 2\zeta\bar{y}' + \bar{y} + \mathrm{sign}(y)\ |\bar{y}|^{\alpha} = \bar{F}\sin(\bar{\omega}\tau), \tag{7.1}$$

where $\alpha \in \{2,\ 2.5,\ 3.5\}$, with the damping ratio set to $\zeta = 5\%$ and the nondimensional forcing amplitude $\bar{F}$ varied across cases.

Case $\alpha = 2$: Helmholtz oscillator

This configuration corresponds to a Helmholtz-type oscillator, characterised by a quadratic stiffness nonlinearity. Figure 7.4 illustrates the predicted fundamental resonance curves at various forcing amplitudes, overlaid with HB continuations depicted as black dashed lines. Table 7.2 presents the corresponding maximum relative amplitude errors.

Forcing amplitude $\bar{F}$ [-]	**1**	**2**	**3**
Max. relative error [%]	4.93	7.31	8.53
Frequency at max. error $\bar{\omega}$ [-]	1.70	1.91	2.06

Table 7.2: Maximum relative amplitude error for $\alpha = 2$ (Helmholtz oscillator)

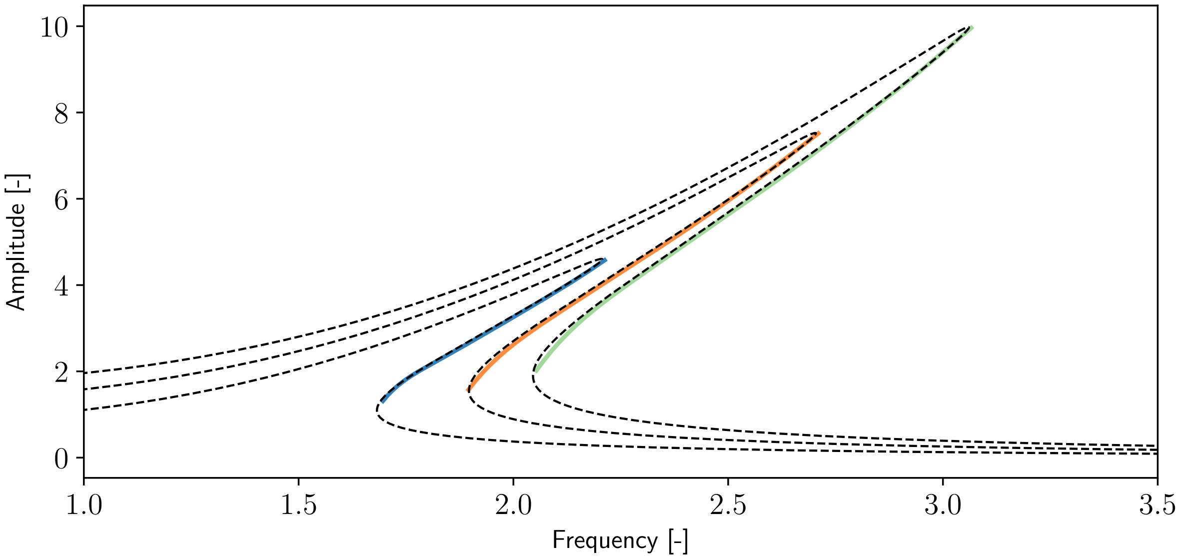


Figure 7.4: Predicted fundamental resonance curves (solid lines) for $\bar{F} = 1$ (—), $\bar{F} = 2$ (—), and $\bar{F} = 3$ (—), compared to the corresponding HB continuations (black dashed lines). Results are shown for a nonlinear system with $\alpha = 2$.

The predictions exhibit good agreement with the reference curves, with the highest discrepancies observed near the lower-amplitude fold. Once again, this region is particularly challenging due to the combination of high curvature and steep gradient of the response, as previously discussed in Chapters 5 and 6.

Case $\alpha = 2.5$

This case corresponds to an oscillator governed by a non-integer exponent, representing intermediate behaviour between quadratic and cubic nonlinearities. As illustrated in Figure 7.5 and Table 7.3, the model maintains high predictive accuracy across all tested forcing amplitudes. Relative errors remain below 3% across all examined forcing amplitudes. The lower error values compared to the $\alpha = 2$ case may be attributed to the presence of Helmholtz-Duffing-type dynamics in the training set, which closely approximates the characteristics of systems with $\alpha = 2.5$.

Forcing amplitude $\bar{F}$ [-]	**1**	**2**	**3**
Max. relative error [%]	1.91	2.65	2.64
Frequency at max. error $\bar{\omega}$ [-]	1.69	1.95	2.14

Table 7.3: Maximum relative amplitude error for $\alpha = 2.5$

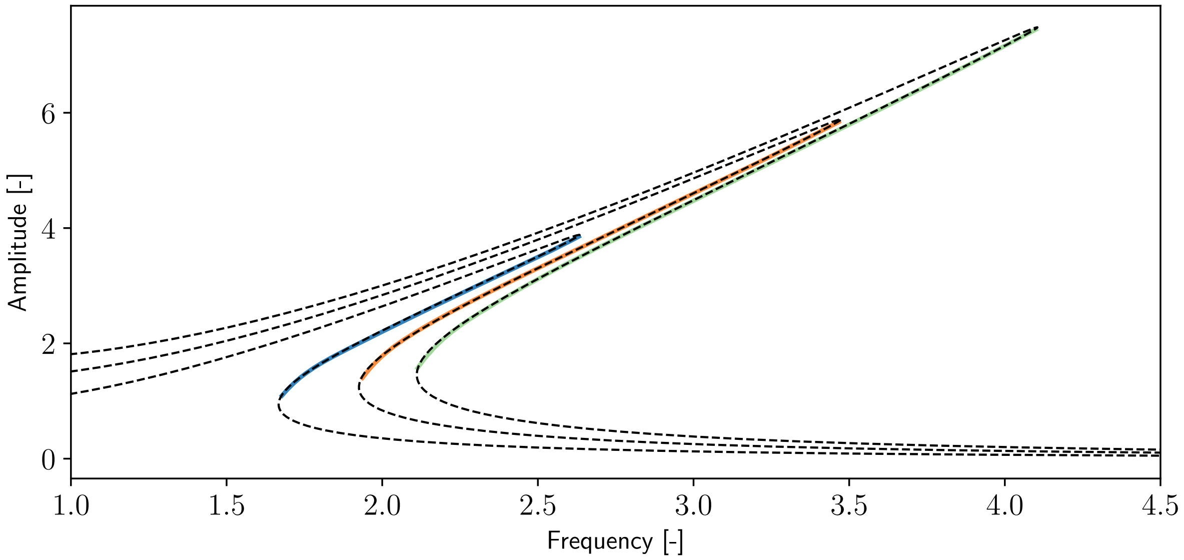


Figure 7.5: Predicted fundamental resonance curves (solid lines) for $\bar{F} = 1$ (—), $\bar{F} = 2$ (—), and $\bar{F} = 3$ (—), compared to the corresponding HB continuations (black dashed lines). Results are shown for a nonlinear system with $\alpha = 2.5$.

These results once again confirm the model's capacity to generalise toward nonlinearities not explicitly encountered during training but lying within the dynamic envelope of the training data.

Case $\alpha = 3.5$

Figure 7.6 and Table 7.4 present the model's predictive performance for the final test case, characterised by $\alpha = 3.5$. Despite the increased nonlinearity, the predictions remain highly accurate across all forcing levels.

Forcing amplitude $\bar{F}$ [-]	**0.5**	**1**	**2**
Max. relative error [%]	3.73	3.13	1.93
Frequency at max. error $\bar{\omega}$ [-]	1.46	1.71	2.00

Table 7.4: Maximum relative amplitude error for α= 3.5

Taken together, the three test cases ($\alpha = 1$, 2.5, and 3.5) demonstrate the model's consistent accuracy across increasing levels of nonlinearity.

This highlights the promise of building sufficiently rich datasets incorporating diverse oscillator types, which, coupled with the model's extrapolation capabilities, could lead to generalisable frameworks for nonlinear resonance prediction. While this chapter focused on SDOF systems and fundamental resonances, future developments should extend the approach to secondary resonances and MDOF sys-

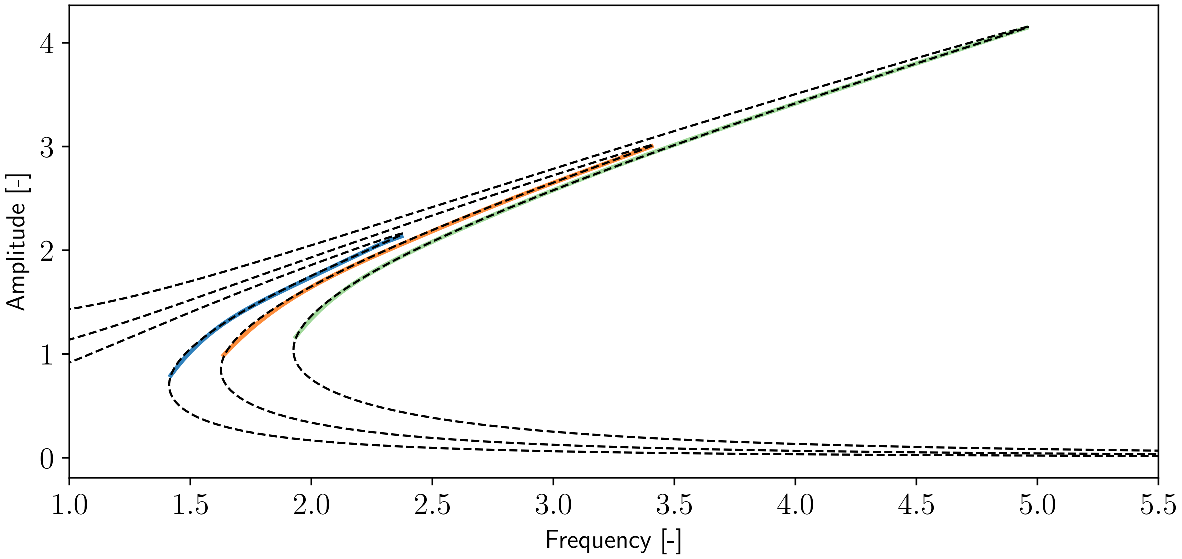


Figure 7.6: Predicted fundamental resonance curves (solid lines) for $\bar{F} = 0.5$ (—), $\bar{F} = 1$ (—), and $\bar{F} = 2$ (—), compared to the corresponding HB continuations (black dashed lines). Results are shown for a nonlinear system with $\alpha = 3.5$.

tems. Such advancements would enable robust, data-driven tools applicable in experimental scenarios where experimental continuation methods may fail due to sensitivity or robustness limitations. Since the only required inputs are the stable branches, readily obtainable through standard experimental modal analysis, this approach remains practically viable for experimental implementation.

7.5 Chapter conclusion and discussion

Chapter conclusion

This chapter assessed the generalisation capabilities of the deep learning model when trained on a heterogeneous dataset comprising frequency response data from three structurally distinct nonlinear oscillators: Duffing, Helmholtz-Duffing, and piecewise linear stiffness systems. The evaluation was conducted through a series of tests that involved both numerical and experimental data.

The first part focused on the Duffing oscillator, using the same test set introduced in Chapter 5. It was shown that expanding the training dataset to include multiple nonlinear systems did not deteriorate performance on the Duffing oscillator. The average F1 score remained unchanged from that of a model trained solely on Duffing data, indicating that increased training diversity does not compromise predictive accuracy.

The second part examined the robustness of the augmented model on experimental data from the electronic Duffing oscillator, as presented in Chapter 6. Predictions remained consistent with experimental measurements, with only marginal variations in relative error. This result confirms that incorporating different nonlinearities during training does not compromise the model's performance on real-world data.

Finally, the generalisation ability of the model was tested on nonlinear systems not present in the training set, characterised by nonlinear restoring forces of the form $\bar{y}^{\alpha}$ with $\alpha \in \{2, 2.5, 3.5\}$. Even in the presence of previously unseen dynamics, the model demonstrated strong predictive accuracy, with relative amplitude errors remaining below 9% in the worst-case scenario. These results suggest that the model is capable of extrapolating its learned representations to unseen types of nonlinear behaviour, provided that they remain within the dynamic envelope implicitly defined by the training data.

Overall, the findings of this chapter support the conclusion that training on a structurally diverse set of nonlinear oscillators enables the model to generalise effectively across a broad spectrum of nonlinear dynamic systems.

CHAPTER 8

Conclusion and perspectives

Conclusion

The primary purpose of this work was to explore how machine learning can be leveraged to support the experimental analysis of nonlinear systems, particularly in the context of identifying nonlinear frequency response curves. Nonlinear systems are characterised by complex phenomena, including multistability, bifurcations, jump discontinuities, and the emergence of secondary resonances, both attached and detached. These phenomena render standard experimental modal analysis methods, such as swept-sine and stepped-sine excitation, insufficient for comprehensive system characterisation. Specifically, unstable solution branches remain inaccessible with these methods.

This master's thesis investigated whether machine learning can overcome this limitation by learning a mapping from standard experimental data to the complete nonlinear response, including unstable and even isolated responses. To the best of the author's knowledge, this is the first research to propose and implement such an approach.

Two central research questions were addressed:

1. Can machine learning approaches bridge the gap between standard experimental model analysis (EMA) testing and nonlinear frequency response?
2. How well does the machine learning model generalise when applied to both experimental data and nonlinear numerical systems not represented in the training set?

To address these questions, the thesis was structured to provide both theoretical background and practical implementation, progressing from foundational principles to numerical and experimental validations.

Chapters 2 and 3 established the theoretical and methodological groundwork. Chapter 2 quantitatively illustrated the discrepancy between linear and nonlinear resonance behaviour using the Duffing oscillator, further enriched with studies on Helmholtz-Duffing and piecewise linear oscillators. It emphasised how nonlinear dynamics give rise to multiple resonances, bifurcations, and complex stability landscapes. Chapter 3 introduced supervised deep learning, with a focus on convolutional neural networks, training principles, loss functions tailored to imbalanced binary classification, and hyperparameter optimisation strategies.

Chapter 4 outlined the methodology for constructing the training dataset, which is a critical component of the machine learning framework developed in this work. Frequency response curves were computed using harmonic balance continuation and subsequently encoded into binary images using a specific grid-based discretisation approach based on Bresenham's algorithm. This image representation allows convolutional neural networks to efficiently process the data and learn the underlying patterns associated with nonlinear dynamic behaviour.

Chapter 5 addressed the first research question by applying the proposed framework to numerical Duffing oscillator simulations. The model trained on fundamental resonance data successfully predicted unstable branches with high accuracy, even under challenging test conditions. An image-based interpolation strategy was introduced to reconstruct continuous curves from grid-based predictions, revealing the trade-offs introduced by discretisation. For secondary resonances, the model also demonstrated the ability to detect subharmonic responses, particularly isolated ones, validating its capacity to generalise across multiple nonlinear phenomena.

Chapter 6 introduced the second research question in the context of experimental validation by transitioning from simulation to real-world data, employing an electronic Duffing oscillator to test the model's performance in experimental conditions. Input data was obtained from swept-sine excitation, while x-ACBC was used to obtain ground-truth FRC. The model predicted the unstable segment of the fundamental resonance with a maximum relative error of only 1.53%. Superharmonic predictions (3:1) showed strong agreement with x-ACBC, in some cases exceeding its resolution. However, the model underperformed in predicting the 2:1 resonance, likely due to structural offsets in the experimental system absent from the training data. Still, the model successfully identified the 1:3 subharmonic response, highlighting its robustness.

Chapter 7 further explored the second research question by evaluating the model's ability to generalise beyond the systems used in training. A convolutional neural network was trained on a composite dataset including Duffing, Helmholtz-Duffing, and piecewise linear stiffness oscillators. Its predictive performance was first assessed on the standard Duffing test set, providing a direct comparison with a baseline model trained solely on Duffing data. The model was then applied to the experimental Duffing oscillator, validating its robustness under real-world measurement conditions. Finally, the evaluation was extended to several additional nonlinear oscillators not present in the training set. In all cases, the model demonstrated strong predictive capabilities, accurately predicting the fundamental resonance of the frequency response curves. These results underscore the model's capacity to learn various representations of nonlinear dynamics and highlight the importance of diverse training data for generalisation.

Limitations

Despite the promising results obtained both numerically and experimentally, several limitations remain:

Firstly, the current method encounters minor discretisation errors caused by the binary image representation of frequency response curves. These errors occur near bifurcation, where rapid changes in amplitude over narrow frequency intervals require high resolution for accurate capture. One straightforward mitigation strategy involves increasing the resolution of the input and output images. A higher resolution directly enhances the frequency and amplitude resolution of the representation. However, this improvement comes at the cost of increased computational and memory demands. Additionally, the resolution cannot be endlessly increased because of hardware limitations. Consequently, while higher resolution can reduce discretisation effects, such limitations cannot be entirely eliminated.

The second limitation arises from the image encoding strategy. The current approach employs Bresenham's line algorithm to convert continuous frequency response curves into binary images. Although this method is computationally efficient and simple to implement, it exhibits reduced accuracy in regions of high curvature or steep slope, particularly near bifurcations. This can lead to unrealistic pixel activations, where an extended vertical segment is assigned to a single frequency bin, resulting in staircasing effects that poorly reflect the actual curve geometry. A potential improvement lies in the use of anti-aliased line rendering techniques, such as Xiaolin Wu's algorithm, which assigns pixel intensities proportionally to their distance from the ideal continuous line. However, this change transforms the task from strict binary segmentation into a regression problem. In practice, this requires replacing the loss with a mean-squared error. This approach produces smoother representations and better preserves the structure of the frequency response. Alternatively, more advanced binary encoding schemes tailored to the specific characteristics of the frequency response curves may be developed to enhance precision without significantly increasing computational complexity.

Lastly, although the model demonstrated strong performance in predicting the resonances observed in the experimental setup, such as the fundamental, the 1:3 subharmonic, and the 3:1 superharmonic, it failed to accurately predict the 2:1 superharmonic resonance branch. As discussed in Chapter 6, this branch becomes stable and observable due to a symmetry-breaking bifurcation induced by a small DC offset in the experimental electronic Duffing oscillator. The presence of this offset transforms an otherwise unstable mathematical solution into a physically realisable resonance, significantly altering the structure of the frequency response curve and leading to dynamic behaviour not captured in the training dataset. As a result, the model was unable to generalise to this particular configuration. This highlights the importance of constructing a diverse and representative dataset that accounts for a broad range of nonlinear responses and system perturbations, thereby enhancing generalisation to real-world conditions.

Perspectives

This master's thesis opens several promising directions for future research in the domain of nonlinear vibrations and data-driven modelling.

First, the present work has primarily focused on SDOF Duffing oscillators. Although Chapter 7 introduced a combined dataset including Helmholtz-Duffing and piecewise-linear stiffness oscillators, many other types of nonlinearities remain unexplored. These include higher-order polynomial stiffness, Coulomb friction, and different non-smooth restoring forces. Extending the current framework to encompass such nonlinearities would be a crucial step toward developing a model capable of generalising to a broader class of nonlinear dynamical systems.

Second, extending the proposed method to MDOF systems constitutes an essential research direction. MDOF systems exhibit modal interactions that couple the response of individual modes, thereby introducing additional complexity into the overall dynamic behaviour. Addressing these interactions would contribute to the development of a more generalised model, capable of capturing the dynamics of a wider class of mechanical structures beyond SDOF systems. In parallel, such advancements would also facilitate the application of the proposed methodology to more complex and experimentally challenging setups.

Third, the dataset generation framework itself could be improved and made more robust. Possible improvements include implementing branch-switching mechanisms to more faithfully capture the complete frequency response curve, including disconnected solution branches. Alternatively, existing continuation toolboxes that incorporate such advanced methods could be leveraged.

Finally, this work demonstrates the effectiveness of using a classical U-Net architecture to predict unstable branches of FRCs, but it also highlights several avenues for further research and improvement. While the current implementation is based on the classical U-Net architecture, variants such as the Attention U-Net [54], Residual U-Net [83], or different encoder-decoder configurations could lead to improved performance by better capturing spatial hierarchies or enhancing gradient flow. Additionally, while the proposed method addresses the problem from a logical and well-founded perspective, exploring alternative paradigms may reveal different or even superior approaches. Therefore, there is still room for improvement, both within the current methodological framework and beyond, through the investigation of fundamentally different deep learning strategies.

In summary, future work should not only extend the methodology to richer nonlinearities and higher-dimensional systems but also explore enhanced network architectures and alternative deep learning paradigms to further improve accuracy and generalisation.

APPENDIX A

Hyperparameter optimisation

A.1 First-stage hyperparameter optimisation

Figure A.1 presents the evolution of the F1 score on both the validation and test sets throughout the training epochs. The different curves correspond to various combinations of focal loss hyperparameters, namely α and γ.

The corresponding F1 scores at the final epoch, along with the optimal threshold τ values, are reported in Table A.1. These results include a comparison with a standard binary cross-entropy baseline.

		$\alpha = 0.6$			$\alpha = 0.7$			$\alpha = 0.8$			$\alpha = 0.9$		
	BCE	$\gamma = 1.0$	$\gamma = 2.0$	$\gamma = 3.0$	$\gamma = 1.0$	$\gamma = 2.0$	$\gamma = 3.0$	$\gamma = 1.0$	$\gamma = 2.0$	$\gamma = 3.0$	$\gamma = 1.0$	$\gamma = 2.0$	$\gamma = 3.0$
Validation set	0.811	0.828	0.824	0.824	0.827	0.826	0.834	0.825	0.835	0.835	0.820	0.819	0.820
Test set	0.827	0.830	0.829	0.827	0.828	0.830	0.835	0.825	0.836	0.835	0.821	0.823	0.821
Optimal τ	0.400	0.500	0.500	0.500	0.550	0.550	0.550	0.600	0.600	0.550	0.750	0.650	0.600

Table A.1: F1 Scores and optimal threshold τ for different α and γ values, including BCE baseline, on the last epoch.

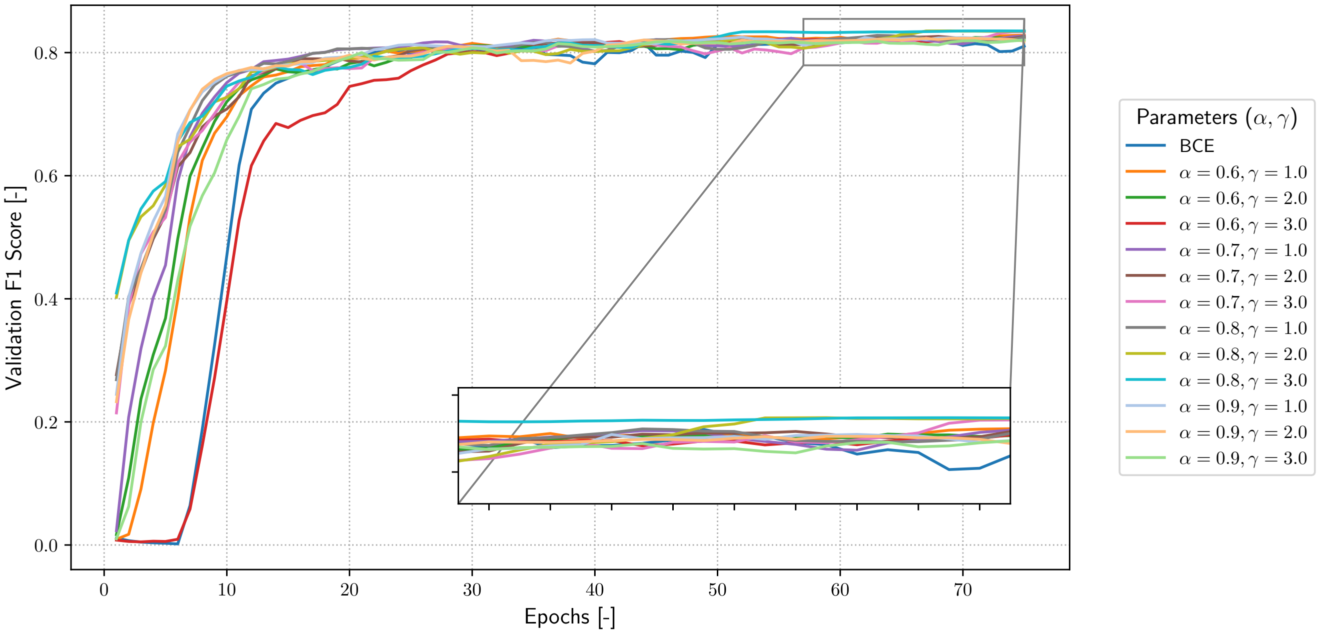
A.2 Second-stage hyperparameter optimisation

A subsequent hyperparameter optimisation was conducted to further enhance the model architecture. This second-stage search varied the network depth, the number of convolutional layers per depth level, the number of filters, the learning rate, and the batch size.

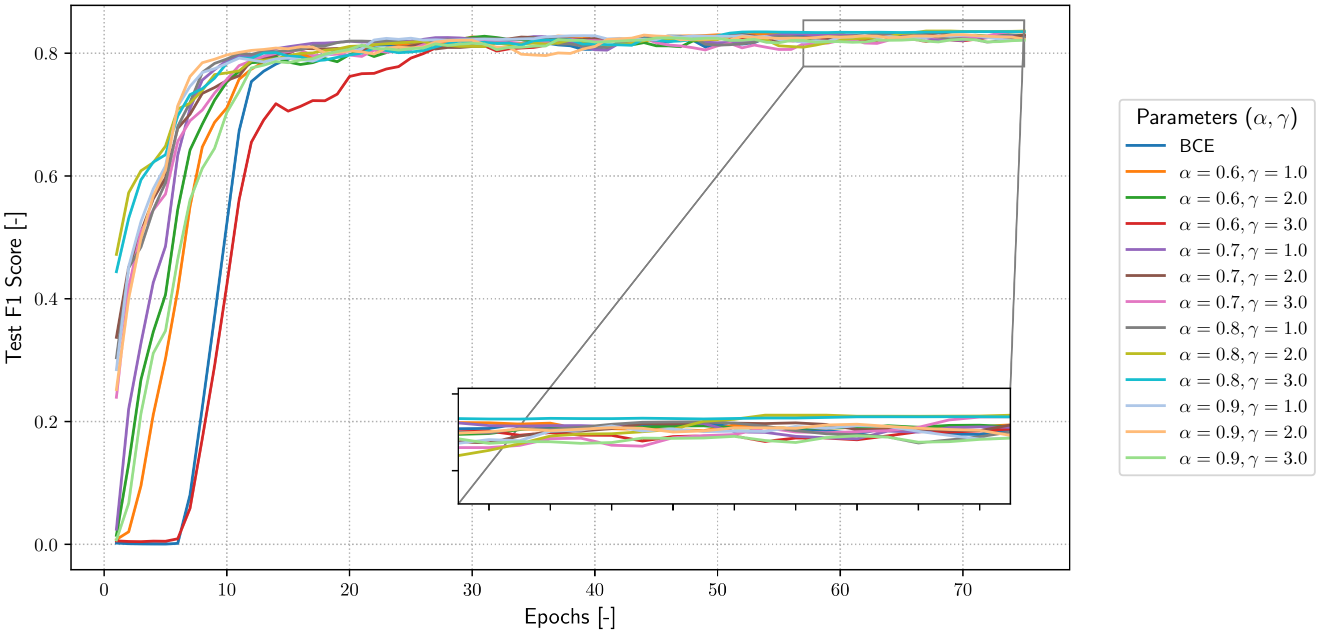
Table A.2 summarizes the validation and test set F1 scores for each architecture explored, including a comparison with the original U-Net configuration proposed by Ronneberger [63].

A.3 Kernel size comparison

Finally, the influence of the convolutional kernel size on model performance was assessed. Figure A.3 compares the F1 score evolution for models trained with kernels of size 3 and 5, using the best hyperparameters obtained from the two-stage optimisation process.

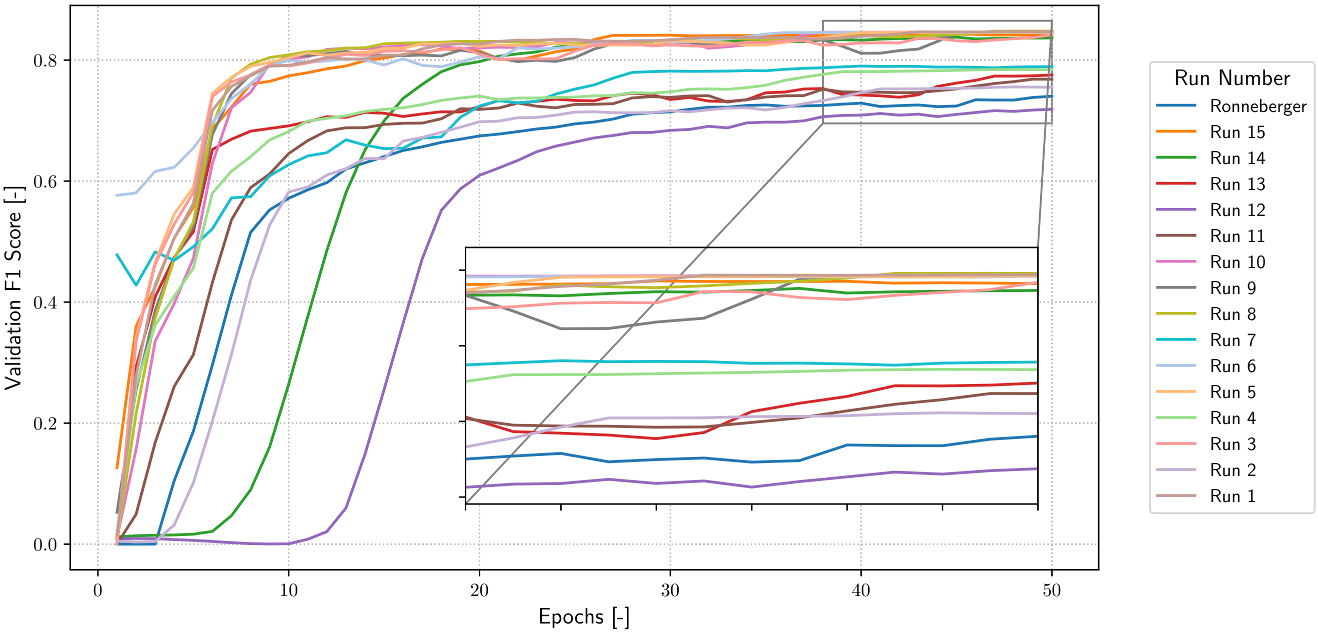


(a) Validation Set

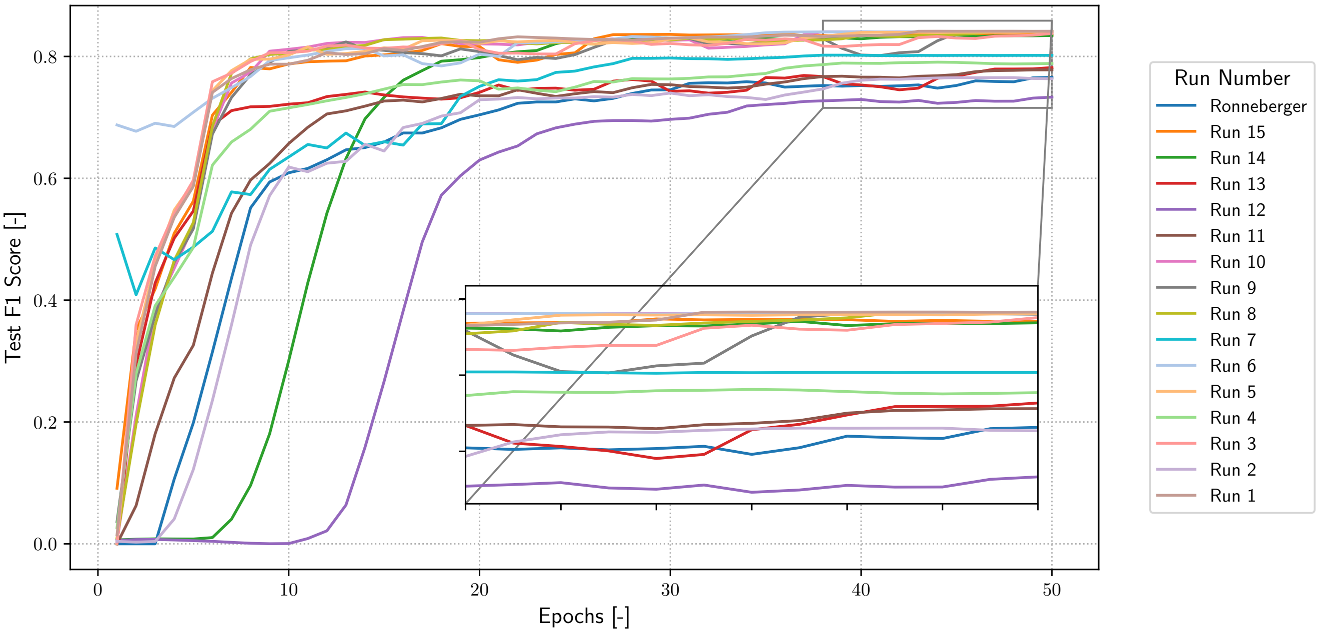


(b) Test Set

Figure A.1: Evolution of the F1 score over training epochs on the validation (a) and test (b) sets for different focal loss hyperparameters α and γ. The model architecture consists of four encoding and decoding layers, three convolutional operations per layer, an initial number of 16 filters, a learning rate of 10^{-4}, and a batch size of 128.



(a) Validation set



(b) Test set

Figure A.2: Evolution of the F1 score over training epochs on the validation (a) and test (b) sets for various model architectures obtained through a random search procedure.

Run	Depth	# Convolutions	Filters	Learning rate	Batch size	Val F1 score	Test F1 score
Ronneberger [63]	4	[2,2,2,2]	64	1e−4	32	0.7402	0.7557
15	4	[2,4,5,5]	32	8.31e−4	32	0.8411	0.8353
14	5	[3,4,5,5,5]	16	1.60e−5	32	0.8366	0.8344
13	3	[2,4,5]	8	5.94e−4	32	0.7754	0.7817
12	3	[2,5,5]	16	1.80e−5	32	0.7187	0.7332
11	3	[4,5,5]	64	1.00e−5	32	0.7684	0.7780
10	4	[2,5,5,5]	16	1.55e−4	32	0.8467	0.8410
9	5	[3,4,5,5,5]	8	7.65e−4	32	0.8460	0.8400
8	5	[3,4,5,5,5]	32	1.15e−4	32	0.8479	0.8411
7	3	[4,5,5]	64	7.10e−4	32	0.7892	0.8018
6	4	[3,4,5,5]	64	2.24e−4	32	0.8461	0.8407
5	5	[3,5,5,5]	8	3.83e−4	32	0.8463	0.8403
4	3	[2,5,5]	16	1.45e−4	32	0.7842	0.7885
3	4	[3,4,5,5]	8	4.78e−4	32	0.8420	0.8378
2	3	[2,5,5]	8	8.00e−5	32	0.7552	0.7634
1	5	[3,4,5,5,5]	32	3.50e−5	32	0.8469	0.8414

Table A.2: Validation and test of the F1 scores for different model configurations.

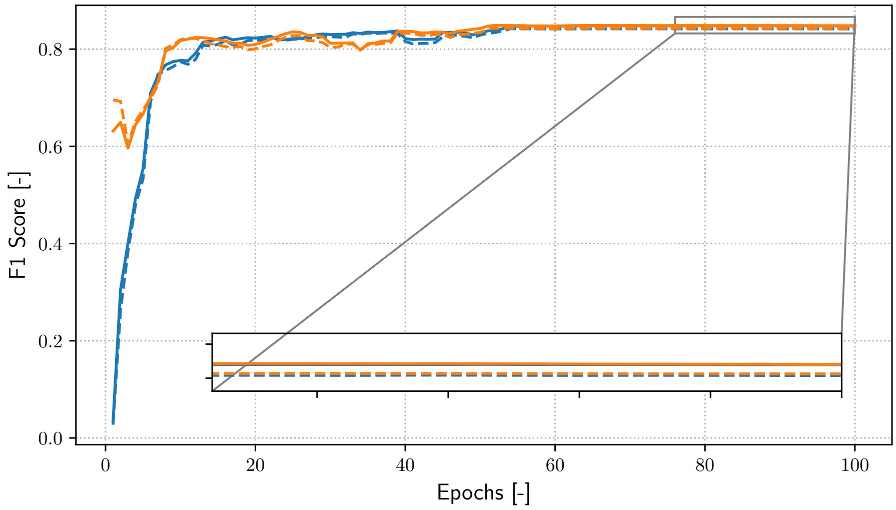


Figure A.3: Evolution of the F1 score on the validation set (solid lines) and test set (dashed lines) for models trained with a kernel size of 3 (blue) and 5 (orange).

Bibliography

[1] Gaëtan Abeloos. *Control-based methods for the identification of nonlinear structures.* PhD thesis, Université de Liège, 2022.

[2] Zaid Ahsan, Harry Dankowicz, Mingwu Li, and Jan Sieber. Methods of continuation and their implementation in the coco software platform with application to delay differential equations. *Nonlinear Dynamics*, 107(4):3181–3243, 2022.

[3] Roberto Alcorta, Sebastien Baguet, Benoit Prabel, Philippe Piteau, and Georges Jacquet-Richardet. Period doubling bifurcation analysis and isolated sub-harmonic resonances in an oscillator with asymmetric clearances. *Nonlinear Dynamics*, 98(4):2939–2960, 2019.

[4] Onur Avci, Osama Abdeljaber, Serkan Kiranyaz, Mohammed Hussein, Moncef Gabbouj, and Daniel J. Inman. 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, 2021. URL: `https://www.sciencedirect.com/science/article/pii/S0888327020304635`, `doi:10.1016/j.ymssp.2020.107077`.

[5] KELLER H. B. Numerical solution of bifurcation and nonlinear eigenvalue problems. *Applications of Bifurcation Theory.*, 1977. URL: `https://cir.nii.ac.jp/crid/1570572699163817472`.

[6] Prabakaran Balasubramanian, Giovanni Ferrari, and Marco Amabili. Identification of the viscoelastic response and nonlinear damping of a rubber plate in nonlinear vibration regime. *Mechanical Systems and Signal Processing*, 111:376–398, 2018. URL: `https://www.sciencedirect.com/science/article/pii/S0888327018301882`, `doi:10.1016/j.ymssp.2018.03.061`.

[7] David A.W. Barton and S.G. Burrow. Numerical continuation in a physical experiment: Investigation of a nonlinear energy harvester. *Journal of Sound and Vibration*, 332(22):6007–6017, 2013. `doi:10.1016/j.jsv.2013.06.032`.

[8] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.

[9] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[10] Bernard Brogliato. *Nonsmooth Mechanics: Models, Dynamics and Control*. Springer, Cham, Switzerland, 3rd edition, 2016. `doi:10.1007/978-3-319-28664-8`.

[11] Gheorghe Bujgoi and Dorin Sendrescu. Tuning of pid controllers using reinforcement learning for nonlinear system control. *Processes*, 13(3), 2025. URL: `https://www.mdpi.com/2227-9717/13/3/735`, `doi:10.3390/pr13030735`.

[12] M. Emre Celebi and Kemal Aydin. *Unsupervised Learning Algorithms*. Springer Publishing Com-

pany, Incorporated, 1st edition, 2016.

[13] Tony F Chan. Newton-like pseudo-arclength methods for computing simple turning points. *SIAM journal on scientific and statistical computing*, 5(1):135–148, 1984.

[14] Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3):88, July 2022. `doi:10.1007/s10915-022-01939-z`.

[15] Filip De Coninck, Wim Desmet, Paul Sas, and David Vaes. Multisine shock and vibration testing using a high-frequency 6-dof shaker table. In *Proceedings of the 10th Int. Congres on Sound and Vibration*, pages 1205–1212, 2003.

[16] Leopoldo de Oliveira, Paulo Varoto, and M.A.S. Peres. Shaker structure interaction: Overview and updated results. *18th International Congress on Sound and Vibration 2011, ICSV 2011*, 3:2516–2523, 01 2011.

[17] Nikolaos Dervilis, Thomas E Simpson, David J Wagg, and Keith Worden. Nonlinear modal analysis via non-parametric machine learning tools. *Strain*, 55(1):e12297, 2019.

[18] T. Detroux, L. Renson, L. Masset, and G. Kerschen. The harmonic balance method for bifurcation analysis of large-scale nonlinear mechanical systems. *Computer Methods in Applied Mechanics and Engineering*, 296:18–38, 2015. URL: `https://www.sciencedirect.com/science/article/pii/S0045782515002297`, `doi:10.1016/j.cma.2015.07.017`.

[19] Thibaut Detroux. *Performance and robustness of nonlinear systems using bifurcation analysis*. PhD thesis, Universite de Liege, 2016.

[20] Annick Dhooge, Willy Govaerts, and Yuri A. Kuznetsov. MATCONT: A MATLAB package for numerical bifurcation analysis of ODEs. *ACM Transactions on Mathematical Software*, 29(2):141–164, 2003. `doi:10.1145/779359.779362`.

[21] Tilan Dossogne, Jean-Philippe Noël, Chiara Grappasonni, Gaëtan Kerschen, Bart Peeters, Jan Debille, Mark Vaes, and Johan Schoukens. Nonlinear ground vibration identification of an f-16 aircraft - part ii understanding nonlinear behaviour in aerospace structures using sine-sweep testing. 06 2015.

[22] Simon Dräger and Jannik Dunkelau. Evaluating the impact of loss function variation in deep learning for classification, 2022. URL: `https://arxiv.org/abs/2210.16003`, `arXiv:2210.16003`.

[23] David J Ewins. *Modal testing: theory, practice and application*. John Wiley & Sons, 2009.

[24] Michel Géradin and Daniel J Rixen. *Mechanical vibrations: theory and application to structural dynamics*. John Wiley & Sons, 2015.

[25] GhislainRaze. Ghislainraze/thhm: Thhm v1.0, January 2025. `doi:10.5281/zenodo.14775562`.

[26] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. URL: `https://www.deeplearningbook.org`.

[27] John Guckenheimer and Philip Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, volume 42 of *Applied Mathematical Sciences*. Springer, 1983. `doi:10.1007/978-1-4612-1140-2`.

[28] Xuekai Guo, Pengfei Lin, Qiulei Wang, and Gang Hu. Intelligent control of structural vibrations based on deep reinforcement learning. *Journal of Infrastructure Intelligence and Resilience*, 4(2):100136, 2025. URL: `https://www.sciencedirect.com/science/article/pii/S2772991524000550`, `doi:10.1016/j.iintel.2024.100136`.

[29] Giuseppe Habib and Ádám Horváth. Fold bifurcation identification through scientific machine learning. *Physica D: Nonlinear Phenomena*, 472:134490, 2025. URL: `https://www.sciencedirect.com/science/article/pii/S0167278924004408`, `doi:10.1016/j.physd.2024.134490`.

[30] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2nd edition, 2009.

[31] Raouf A. Ibrahim. *Vibro-Impact Dynamics: Modeling, Mapping and Applications*. Springer, Berlin, 2009. `doi:10.1007/978-3-642-00275-5`.

[32] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456, 2015.

[33] Zhenzhen Jin, Deqiang He, Zhenpeng Lao, Zexian Wei, Xianhui Yin, and Weifeng Yang. Early intelligent fault diagnosis of rotating machinery based on iwoa-vmd and dmkelm. *Nonlinear Dynamics*, 111(6):5287–5306, March 2023. `doi:10.1007/s11071-022-08109-8`.

[34] Harrish Joseph, Giuseppe Quaranta, Biagio Carboni, and Walter Lacarbonara. Deep learning architectures for data-driven damage detection in nonlinear dynamic systems under random vibrations. *Nonlinear Dynamics*, 112(23):20611–20636, December 2024. `doi:10.1007/s11071-024-10270-1`.

[35] Taylan Karağaçlı and H. Nevzat Özgüven. Experimental modal analysis of nonlinear systems by using response-controlled stepped-sine testing. *Mechanical Systems and Signal Processing*, 146:107023, 2021. `doi:10.1016/j.ymssp.2020.107023`.

[36] Jinki Kim, R. L. Harne, and K. W. Wang. Predicting non-stationary and stochastic activation of saddle-node bifurcation. *Journal of Computational and Nonlinear Dynamics*, 12(1):011009, 09 2016. `doi:10.1115/1.4034128`.

[37] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[38] Ivana Kovacic and Michael J. Brennan. *The Duffing Equation: Nonlinear Oscillators and Their Behaviour*. John Wiley & Sons, Chichester, UK, 2011. `doi:10.1002/9780470977859`.

[39] Malte Krack and Johann Gross. *Harmonic balance for nonlinear vibration problems*, volume 1. Springer, 2019.

[40] Malte Krack and Johann Gross. *Harmonic Balance for Nonlinear Vibration Problems*. Mathematical Engineering. Springer Nature Switzerland AG, 2019. `doi:10.1007/978-3-030-14023-6`.

[41] Tido Kubatschek and Alwin Förster. Investigation of existing and new approaches to step size control in a continuation framework. *Computers Structures*, 313:107747, 2025. URL: `https://www.sciencedirect.com/science/article/pii/S0045794925001051`, `doi:10.1016/j.compstruc.2025.107747`.

[42] Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. Springer, 2013.

[43] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[44] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. Hyperband: A novel bandit-based approach to hyperparameter optimization. In *International Conference on Learning Representations*, 2017.

[45] Tien-Yien Li. Numerical solution of multivariate polynomial systems by homotopy continuation methods. *Acta numerica*, 6:399–436, 1997.

[46] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[47] JM Liu, YF Xu, and WD Zhu. New coherence function with measurements of one sampling period. *Mechanical Systems and Signal Processing*, 88:354–376, 2017.

[48] A. Motallebi, S. Irani, and S. Sazesh. Analysis on jump and bifurcation phenomena in the forced vibration of nonlinear cantilever beam using hbm. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 38(2):515–524, 2016. `doi:10.1007/s40430-015-0352-2`.

[49] James A Murdock. *Perturbations: theory and methods*. SIAM, 1999.

[50] Ali H Nayfeh. *Introduction to perturbation techniques*. John Wiley & Sons, 2024.

[51] Ali H. Nayfeh and Dean T. Mook. *Nonlinear Oscillations*. Wiley-Interscience, New York, 1995.

[52] NOLISYS. Ni2d softwarre. `https://nolisys.com/`. Accessed: 2025-05-27.

[53] F. Nucera, A. F. Vakakis, D. M. McFarland, L. A. Bergman, and G. Kerschen. Targeted energy transfers in vibro-impact oscillators for seismic mitigation. *Nonlinear Dynamics*, 50(3):651–677, 2007. `doi:10.1007/s11071-006-9189-7`.

[54] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[55] M. Oregui, Z. Li, and R. Dollevoet. An investigation into the modeling of railway fastening. *International Journal of Mechanical Sciences*, 92:1–11, 2015. URL: `https://www.sciencedirect.com/science/article/pii/S0020740314003932`, `doi:10.1016/j.ijmecsci.2014.11.019`.

[56] Benjamin R. Pacini, Robert J. Kuether, and Daniel R. Roettgen. Shaker-structure interaction modeling and analysis for nonlinear force appropriation testing. *Mechanical Systems and Signal Processing*, 162:108000, 2022. URL: `https://www.sciencedirect.com/science/article/pii/S0888327021003952`, `doi:10.1016/j.ymssp.2021.108000`.

[57] Rik Pintelon. *System Identification: A Frequency Domain Approach*. 02 2001. `doi:10.1002/0471723134`.

[58] Simon J.D. Prince. *Understanding Deep Learning*. The MIT Press, 2023. URL: `http://udlbook.com`.

[59] E. Ramm. Strategies for tracing the nonlinear response near limit points. In W. Wunderlich, E. Stein, and K.-J. Bathe, editors, *Nonlinear Finite Element Analysis in Structural Mechanics*, pages 63–89, Berlin, Heidelberg, 1981. Springer Berlin Heidelberg.

[60] Ghislain Raze. An electronic duffing oscillator. 2024. URL: `https://github.com/GhislainRaze/Electronic-Duffing`.

[61] Ghislain Raze and Gaëtan Kerschen. A two-harmonic homotopy method to connect a primary resonance to its secondary resonances. *Proceedings A*, 481, 05 2025. `doi:10.1098/rspa.2024.0875`.

[62] Ludovic Renson, Gaëtan Kerschen, and Bruno Cochelin. Numerical computation of nonlinear normal modes in mechanical engineering. *Journal of Sound and Vibration*, 364:177–206, 2016.

[63] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[64] Abdolvahhab Rostamijavanani, Shanwu Li, and Yongchao Yang. Data-driven nonlinear modal analysis with physics-constrained deep learning: Numerical and experimental study. *arXiv preprint arXiv:2503.08952*, 2025.

[65] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.

[66] R. Seydel. *Practical Bifurcation and Stability Analysis*. Interdisciplinary Applied Mathematics. Springer, 3rd edition, 2009. See Algorithm 5.11, Section 5.6.2.

[67] Steven W. Shaw and Philip J. Holmes. A periodically forced piecewise linear oscillator. *Journal of Sound and Vibration*, 90(1):129–155, 1983. `doi:10.1016/0022-460X(83)90407-8`.

[68] Jan Sieber and Bernd Krauskopf. Control-based continuation of periodic orbits. *SIAM Journal on Applied Dynamical Systems*, 7(2):459–483, 2008. `doi:10.1137/070710145`.

[69] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. Practical bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, pages 2951–2959, 2012.

[70] I.J. SOKOLOV and V.I. BABITSKY. Phase control of self-sustained vibration. *Journal of Sound and Vibration*, 248(4):725–744, 2001. URL: `https://www.sciencedirect.com/science/article/pii/S0022460X01938106`, `doi:10.1006/jsvi.2001.3810`.

[71] Alexandre Spits. Derivative-less control-based nonlinear vibration testing. Master's thesis, University of Liège, 2024.

[72] Alexandre Spits, Ghislain Raze, and Gaëtan Kerschen. Derivative-free arclength control-based continuation for secondary resonances identification. In *IMAC 2025 Proceedings*. River publishers, 11 February 2025.

[73] JJ Stoker. Nonlinear vibrations in mechanical and electrical systems. interscience publ. *Inc., c*, 1950.

[74] Chen Sun, Abhinav Shrivastava, Saurabh Singh, and Abhinav Gupta. Revisiting unreasonable effectiveness of data in deep learning era, 2017. URL: `https://arxiv.org/abs/1707.02968`, `arXiv:1707.02968`.

[75] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2 edition, 2018. URL: `https://mitpress.mit.edu/9780262039246/reinforcement-learning-second-edition/`.

[76] Jakub M. Tomczak. *Deep Generative Modeling*. Springer, 2022. URL: `https://link.springer.com/book/10.1007/978-3-030-93158-2`, `doi:10.1007/978-3-030-93158-2`.

[77] Alexander F Vakakis, Leonid I Manevitch, Yuri V Mikhlin, Valery N Pilipchuk, and Alexandr A Zevin. Normal modes and localization in nonlinear systems. 2001.

[78] Martin Volvert and Gaëtan Kerschen. Phase resonance nonlinear modes of mechanical systems. *Journal of Sound and Vibration*, 511:116355, 2021. URL: `https://www.sciencedirect.com/science/article/pii/S0022460X21004119`, `doi:10.1016/j.jsv.2021.116355`.

[79] Steven Euijong Whang, Yuji Roh, Hwanjun Song, and Jae-Gil Lee. Data collection and quality challenges in deep learning: A data-centric ai perspective, 2022. URL: `https://arxiv.org/abs/2112.06409`, `arXiv:2112.06409`.

[80] K. Worden and P.L. Green. A machine learning approach to nonlinear modal analysis. *Mechanical Systems and Signal Processing*, 84:34–53, 2017. Recent advances in nonlinear system identification. URL: `https://www.sciencedirect.com/science/article/pii/S0888327016300681`, `doi:10.1016/j.ymssp.2016.04.029`.

[81] M.-C. Yu, X. Gao, and Q. Chen. Nonlinear frequency response analysis and jump avoidance design of molecular spring isolator. *Journal of Mechanics*, 32(5):527–538, 2016. `doi:10.1017/jmech.2016.52`.

[82] Yi-Ang Zhang and Songye Zhu. Active nonlinear vibration control of a buckled beam based on deep reinforcement learning. *Journal of Vibration and Control*, 31(1-2):70–81, 2025. `arXiv: https://doi.org/10.1177/10775463241264112`, `doi:10.1177/10775463241264112`.

[83] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.

[84] Tong Zhou and Gaëtan Kerschen. Identification of secondary resonances of nonlinear systems using phase-locked loop testing. Technical report, SSRN Electronic Journal, 2019. Available at SSRN: `https://ssrn.com/abstract=4695306` or `http://dx.doi.org/10.2139/ssrn.4695306`. `doi:10.2139/ssrn.4695306`.

[85] Tong Zhou and Gaëtan Kerschen. Identification of secondary resonances of nonlinear systems using phase-locked loop testing. *Journal of Sound and Vibration*, 590:118549, 2024. URL: `https://www.sciencedirect.com/science/article/pii/S0022460X24003122`, `doi:10.1016/j.jsv.2024.118549`.

[86] Xiangxin Zhu, Carl Vondrick, Charless C. Fowlkes, and Deva Ramanan. Do we need more training data? *International Journal of Computer Vision*, 119(1):76–92, March 2015. URL: `http://dx.doi.org/10.1007/s11263-015-0812-2`, `doi:10.1007/s11263-015-0812-2`.