big data and
cognitive computing

Article

Machine Learning Based Impact Sensing Using Piezoelectric Sensors: From Simulated Training Data to Zero-Shot Experimental Application

Petros Gkertzos [1,*], Johannes Gerritzen [2], Constantinos Tsakonas [3], Stefanos H. Panagiotou [3], Athanasios Kotzakolios [1], Ioannis Katsidimas [1], Andreas Hornig [2,4,5], Siavash Ghiasvand [4], Maik Gude [2], Vassilis Kostopoulos [1,*] and Sotiris Nikoletseas [3]

[1] Applied Mechanics & Vibrations Laboratory, Department of Mechanical Engineering and Aeronautics, University of Patras, Rio Campus, 26500 Patras, Greece; kotzakol@upatras.gr (A.K.)
[2] Institute of Lightweight Engineering and Polymer Technology (ILK), TUD Dresden University of Technology, Holbeinstraße 3, 01307 Dresden, Germany; johannes.gerritzen@tu-dresden.de (J.G.); andreas.hornig@tu-dresden.de (A.H.); maik.gude@tu-dresden.de (M.G.)
[3] Computer Engineering and Informatics Department, University of Patras, 26504 Patras, Greece; spanagiotou@ceid.upatras.gr (S.H.P.); nikole@cti.gr (S.N.)
[4] Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig (ScaDS.AI), TUD Dresden University of Technology, Strehlener Str. 14, 01069 Dresden, Germany; siavash.ghiasvand@tu-dresden.de
[5] Department of Engineering Science, University of Oxford Parks Road, Oxford OX1 3PJ, UK
* Correspondence: up1047297@ac.upatras.gr (P.G.); kostopoulos@upatras.gr (V.K.)

Abstract

Modern impact monitoring systems combine multiple inputs with machine learning (ML) models for impact detection, localization, and event assessment. Their accuracy relies on large, event-representative datasets, used for algorithmic development and ML model training. High-fidelity numerical models can provide augmented datasets by overcoming the cost and time limitations of experimental methods. This research presents an end-to-end numerical methodology for impact detection based on simulation (training) and experimental (testing) data. Initially, a finite element model (FEM) of our experimental setup utilizing piezoelectric transducer (PZT) sensors mounted on a thermoplastic plate is created. From the experimental impact signals, a few consistent cases are identified for feature extraction. A design of experiments explores the range of each parameter, and through surrogate optimization, the material and piezoelectric properties of the setup are determined. Subsequently, a virtual dataset, involving multiple impact cases, is created to train the ML models performing impact detection. Testing with experimental data shows results consistent with literature studies that used only experimental data for both training and testing. This work provides a systematic methodology for representative dataset generation and impact monitoring through ML, while addressing accurate FEM parameter identification from a few experimental tries.

Keywords: finite element model (FEM); impact sensing; piezoelectric transducer (PZT); parameter fitting; surrogate based optimization; machine learning

Check for updates

Academic Editor: Fabrizio Messina

Received: 28 September 2025
Revised: 6 December 2025
Accepted: 17 December 2025
Published: 23 December 2025

1. Introduction

Structural health monitoring (SHM) has gained significant attention, as evidenced by an abundance of publications over the past ten years [1–4]. SHM studies provide an algorithmic, sensing, and data processing perspective to evaluate the condition of the examined structure and, in turn, lead to early fault detection [5–7]. Modern impact monitoring systems have evolved to combine multiple inputs with heuristic algorithms or ML models, capable of impact detection, localization, and event assessment [8–11]. Capineri L. et al. [10] investigate impact localization by analysis of Lamb waves, produced in impact events, which are captured using PZT sensors. Their analysis is based on the discrimination of symmetrical S_0 and antisymmetrical A_0 modes, to extract differences in times of arrival (DToAs) and signal energy values. These features are utilized by a multichannel Analog Front-End Module to assess impact detection. Guo Z. et al. [12] present a multi-channel and multi-function SHM system intended for aerospace structures. The system can perform impact detection and damage quantification by using a set of PZT sensors. Impact monitoring is performed based on average voltage deviation to discover the sensor that is closest to the impact point. Dipietrangelo F. et al. [13] propose an ML approach, based on artificial neural networks (ANNs), capable of low-velocity impact localization on an aluminum plate. They present an implementation of a vibration acquisition system via Pimoroni HAT Explorer Pro and Raspberry Pi, capable of operating with a typical USB power supply of just 5V.

The accuracy and robustness of such systems depend on the size of event representative datasets, used for algorithm development and ML model training, as well as the size of the monitored area. Figure 1 gives a comparative literature analysis of different impact localization works [14–22]. Average localization error (distance between predicted and actual location in mm) is given for monitored areas ranging between 230 mm * 230 mm and 600 mm * 900 mm. Especially at larger surfaces, dataset size significantly affects localization accuracy, while the threshold of 150 tests is identified as an approximate requirement for good results.

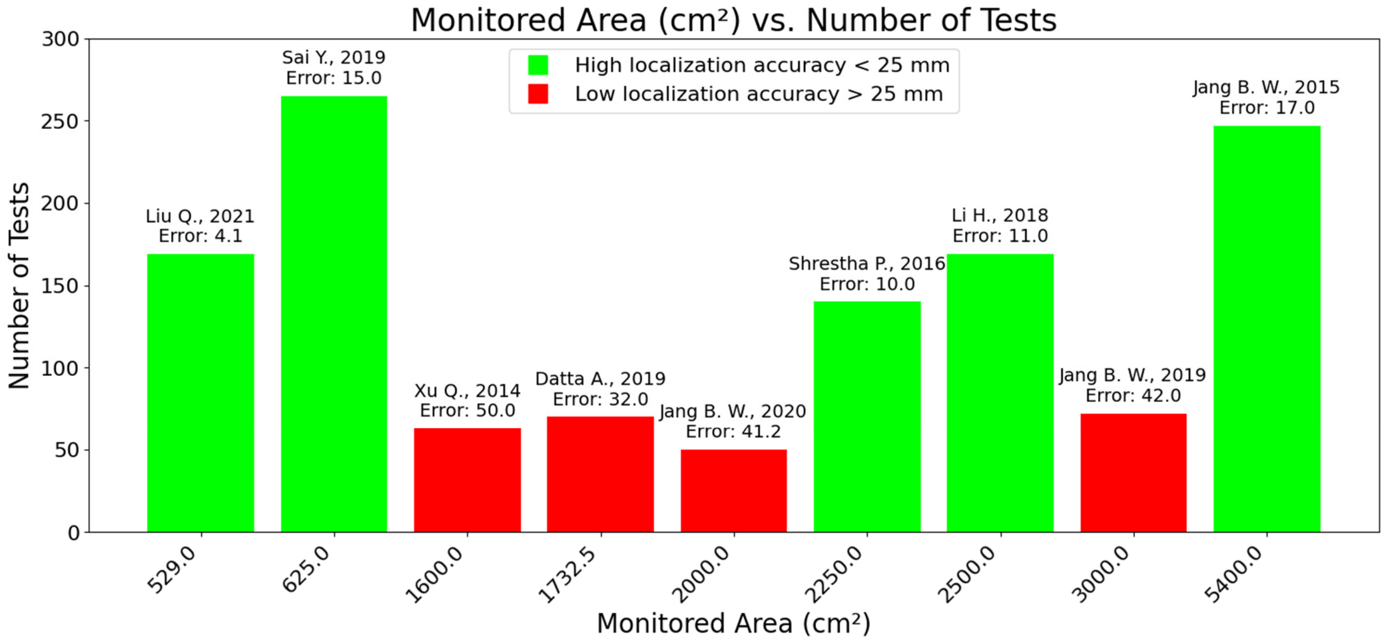


Figure 1. Comparative literature analysis of impact localization works [14–22]. Studies include isotropic and orthotropic panels, all with four sensors near the corners. Despite material differences, the dataset size requirement remains a consistent observation.

High-fidelity numerical models can be employed to augment datasets, thereby overcoming the cost and time limitation of extensive experimental testing [23–25]. The modeling approach and input parameters, both of the constitutive models and purely numerical ones, significantly affect the outputs and thus the reliability of the numerical framework [25,26].

As such, accurate parameter identification from a few experimental tries, without extensive testing of all the system components, can significantly aid in the task of representative (numerical) dataset augmentation.

Hesser et al. [27] used ABAQUS software to simulate the impact of a steel ball on an aluminum plate (150 mm * 150 mm) mounted with four PZT sensors. The numerical model captures the piezoelectric effect by incorporating the piezoelectric coefficient and permittivity matrices in the constitutive model of the PZT. The resulting numerical dataset is used to train impact localization models (ANN and support vector machine (SVM)). Experimental data are used to test these models, revealing that the ANN model provides significantly better prediction performance. Ren L. et al. [28] propose a method for impact localization in composite structures using an adaptively calibrated piezoelectric sensor array. They analyze signals from both FEM simulations and experiments on a carbon-fiber composite panel with a mounted PZT sensor array. These signals are represented by an error calibration matrix with unknown gains and phases. To determine the sensor array error parameters and accurately locate the impact, the 2D-MUSIC algorithm is applied. The cost function is minimized through an adaptive iterative process, improving localization accuracy.

In all the above works, impact localization is performed in mostly experimental or, in some cases, simulation frameworks. The few that utilize hybrid datasets have known numerical input parameters, which were obtained from testing of each system component. To further facilitate the application of existing augmentation techniques, this work presents a methodology for identifying governing parameters of an impact localization setup utilizing PZTs mounted on a thermoplastic plate as sensors. The proposed approach relies solely on tests performed on the impact system. No parameter calibration on the coupon level is required. Features are extracted from the signals captured throughout only four representative experimental impact events. Based on these signals, a two-stage surrogate-based optimization scheme is proposed, capable of efficiently identifying parameter values for both involved materials and the entire impact system. Leveraging the identified parameters, the simulation model is used to create a purely virtual dataset comprising a total of 490 impact events at different locations and from different heights. The virtual dataset is used to train several ML models, predicting impact location from sensor signals. The pretrained models are applied without any adaptation to previously unseen experimental impact events, exhibiting acceptable accuracy. This novel zero-shot application of a model trained on virtual data to real-world scenarios is a key step for the development and implementation of ML-based SHM systems, since obtaining experimental datasets is typically prohibitively expensive.

2. Experimental Data Acquisition

In this section, we summarize the methodology that was followed to create the experimental impact event dataset [29], publicly available at the Zenodo repository [30].

2.1. Experimental Setup

The configuration consists of a thin polymethyl methacrylate (PMMA) plate P with a size of 320 × 317 mm and a thickness of 4.2 mm, featuring four PZTs, each attached at one corner. A steel ball (9.5 mm diameter, 3.53 g weight) is used as an impactor, and the ceramic PZTs (CEB-35D26, Mouser Electronics, Mansfield, TX, USA) [31] act as sensors. To capture the transmitted waves and determine the impact location, the PZT sensors are attached to the plate and interfaced with a microcontroller (Arduino NANO 33 BLE). A laptop is used to log the data that the Arduino collects from the sensors. The setup is illustrated in Figure 2. In this Figure, the plate's surface is divided into a 5 × 5 grid of equal rectangles, which supported impact localization analysis through classification in our previous work [29]. The achieved accuracy, which is a metric that quantifies the proportion of correctly predicted impact regions (rectangles) relative to all tested cases, was higher than 90%, serving as an initial reliability indicator of the experimental dataset.

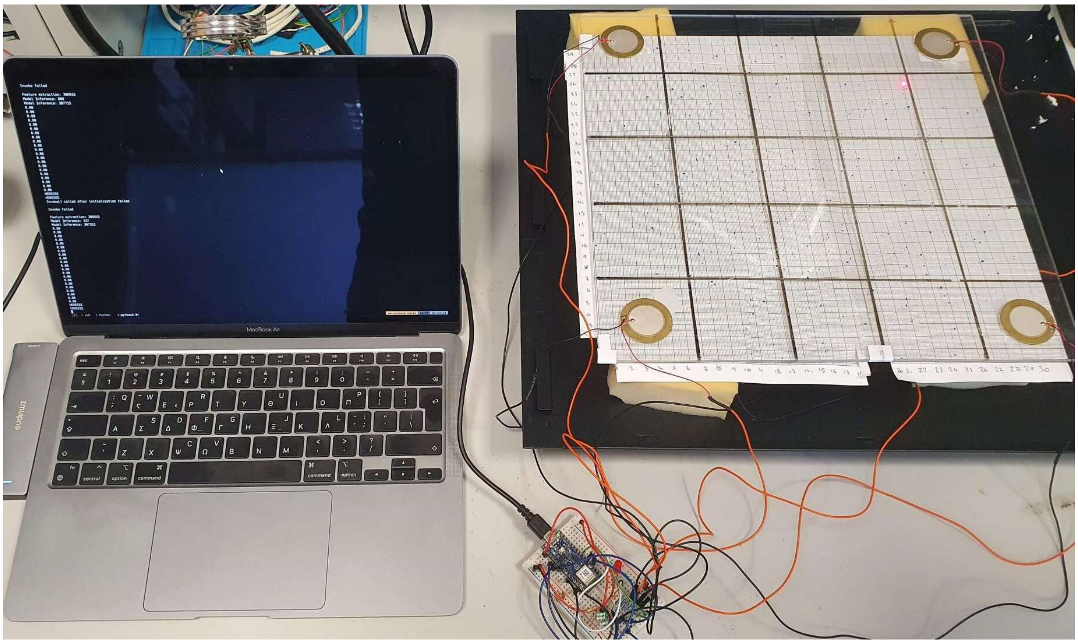


Figure 2. Image of the experimental plate setup.

2.2. Design and Execution of Experiments

Incorporating a systematic design of experiment (DOE) approach is essential for enhancing the quality of the dataset and uncovering the system dynamics. By varying the x, y coordinates, and drop heights, the experiments are designed to capture a comprehensive range of impact scenarios. This deliberate variation, given enough sampling, denotes an ample dataset while providing deeper insight into the underlying system dynamics and the identification of how different impact parameters affect each sensor signal. In this study, the Sobol algorithm is employed as a space-filling DOE generator. Sobol sequences, also referred to as LP_T sequences, produce a quasi-random, low-discrepancy sequence where test points are distributed within the boundaries of the design space. The boundaries of the design are user-defined for each input parameter. As such, the x and y coordinates are set within the plate dimensions, while eliminating coordinates on top or directly near the sensors to prevent electronics damage or localized interference that could compromise the sensor signals. The impact height is set within 100 and 200 mm to allow for elastic impacts with zero inflicted damage on the plate, thus not altering the wave propagation characteristics. The most notable feature of Sobol sequences is their ability to scatter test points, effectively eliminating subjective bias and allowing good sampling of the configuration space [32–34]. Figure 3 shows a 2D illustration of the combinations of impact coordinates and drop heights based on Sobol sampling.

The steel ball drop setup used to perform the above Sobol-generated impact tests is depicted in Figure 4. Specifically, the impact is generated by releasing a steel ball B from a varying height of 100…200 mm, with increments of 5 mm, onto the plate P. The ball is released into free fall (with zero initial velocity assisted by the Drop driver). The Laser device and Drop driver are fixed, and the impact location is modified by repositioning the plate to the desired x and y coordinates, verified using the Laser beam. The experiments were conducted in laboratory conditions, while the existence of a low-density foam minimizes the influence of possible environmental vibrations.

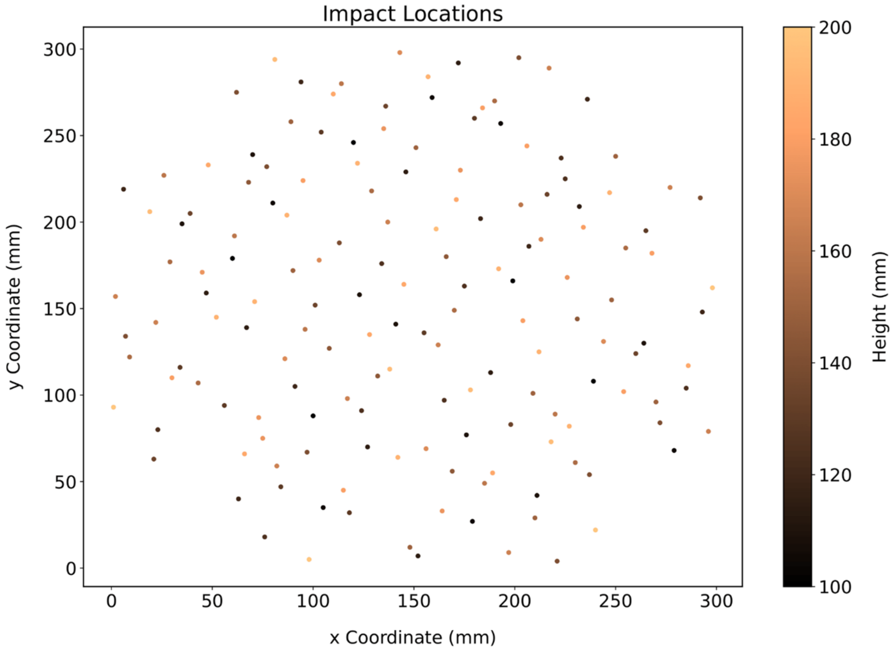


Figure 3. Two-dimensional representation of impact coordinates and fall height combinations (with different colors) based on Sobol sampling. In total, the dataset consists of 771 drop tests (multiple repetitions of 159 distinct impact cases).

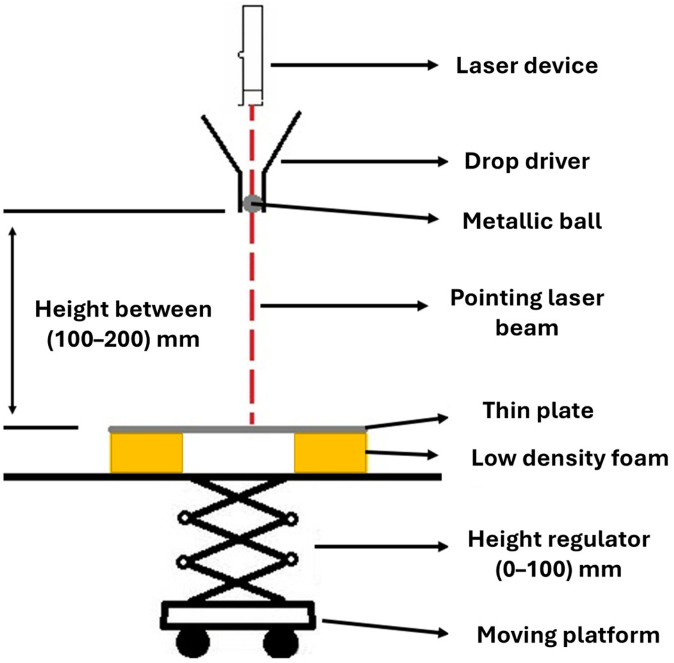


Figure 4. Schematic representation of the steel-ball drop setup used to generate the experimental impact dataset.

Each impact event is recorded over a total duration of 50 ms, using a high sampling frequency of 100 kHz, resulting in 5000 samples per case. Although the impact signals typically last for about 500 samples (5 ms), the extended recording window ensures that both the initial transient and the full system response are captured. This sampling strategy enhances the temporal resolution and fidelity of the dataset, enabling detailed analysis of the system's dynamic behavior.

3. Numerical Modeling

Figure 5 presents the corresponding FEM created in LS-DYNA [35]. The PMMA plate is instrumented with four PZT sensors at the corners, positioned 30 and 33 mm from each adjacent edge. The impact location is indicated by a ball bearing, with its coordinates defined as x_{impact}, y_{impact}. The red dashed square outlines the inner region boundary, where the majority of impact events are evident ($60 \leq x, y \leq 260$ mm). To decrease the computational cost, 101,440 4-node quadrilateral shell elements are used for the plate, and 632 8-node hexahedral solid elements are used for each PZT. To ensure numerical stability and accurately capture wave propagation phenomena, the integration time step Δt and the finite element size L_e must comply with the following equations [36]:

$$L_e \leq \frac{\lambda_{min}}{10} \quad (1)$$

$$\Delta t \leq \frac{1}{20 * f_{max}} \quad (2)$$

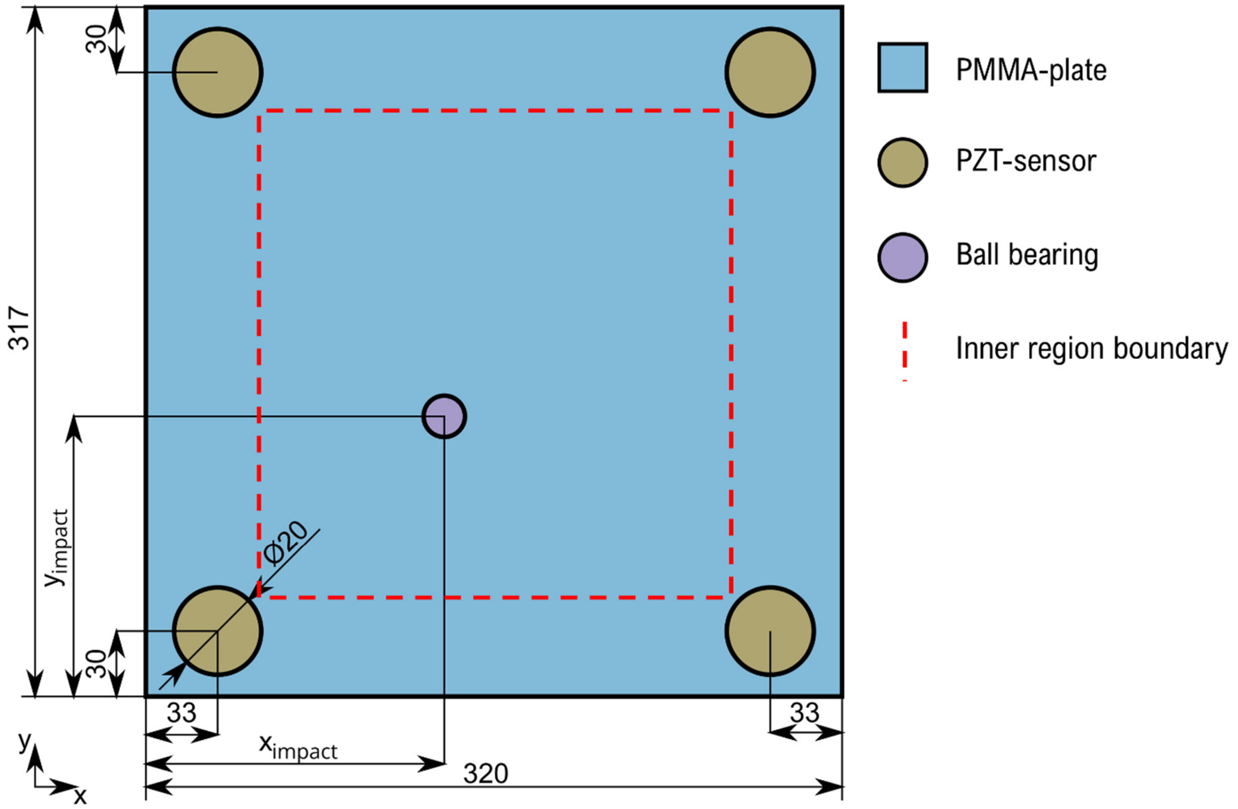


Figure 5. Dimensions of investigated impact system.

Table 1 presents the measured dimensions and properties of the structure. Young's modulus will be determined by the proposed two-step optimization procedure.

Due to the low energy of the impact, the deformation remains in the linearly elastic region. Hence, Hooke's law is used to model all materials. In this work, the PMMA plate is considered isotropic, and PZT sensors are orthotropic.

In piezoelectric materials, the electric displacement field $\boldsymbol{D}$ depends on both the applied electric field and the material's polarization [37,38]. This relationship is generally expressed as

$$\boldsymbol{D} = \boldsymbol{e} * \boldsymbol{E_f} + \boldsymbol{P} \quad (3)$$

where $\boldsymbol{e}$ is the material permittivity, $\boldsymbol{E_f}$ is the electric field vector, and $\boldsymbol{P}$ is the polarization vector.

Table 1. Properties of investigated objects.

Property	Symbol	Unit	Plate	Ball	PZT
Material	-	-	PMMA	Steel	Ceramic
Dimension	$l \times b$	mm × mm	320 × 317	-	-
Diameter	D	mm	-	9.5	20
Thickness	t	mm	4.2	-	0.25
Density	ϱ	kg/(m^3)	1157	7862	N/A
Weight	m	g	493	3.5	N/A

The permittivity of the material can be written as

$$\boldsymbol{e} = e_0 * \boldsymbol{e_r} = e_0 * (1 + \boldsymbol{\chi_e}) \tag{4}$$

where e_0 is the permittivity of free space, $\boldsymbol{e_r}$ is the relative permittivity tensor, and $\boldsymbol{\chi_e}$ is the electric susceptibility tensor of the medium.

In piezoelectric materials, the polarization $\boldsymbol{P}$ is directly proportional to the mechanical stress applied, as described by the following equation [37,38]:

$$\boldsymbol{P} = \boldsymbol{d} * \boldsymbol{\sigma} \tag{5}$$

where $\boldsymbol{d}$ is the piezoelectric coefficient tensor and $\boldsymbol{\sigma}$ is the applied stress.

The general constitutive relation for the coupled electromechanical behavior of piezoelectric materials is given by the following [37,38]:

$$\begin{bmatrix} \boldsymbol{\varepsilon} \\ \boldsymbol{D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{s} & \boldsymbol{d}^T \\ \boldsymbol{d} & \boldsymbol{e} \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma} \\ \boldsymbol{E_f} \end{bmatrix} \tag{6}$$

where $\boldsymbol{\varepsilon}$ is the mechanical strain tensor, and $\boldsymbol{s}$ is the elastic compliance matrix. Within the scope of this work, the PZTs are in sensing mode (not connected to a voltage source in order to act as an actuator), which corresponds to open-circuit electrical boundary conditions [39,40]. In this case, the electric displacement $\boldsymbol{D}$ remains constant (approximately zero net charge flow), and an internal electric field $\boldsymbol{E_f}$ develops in response to mechanical stress, resulting in a measurable voltage at the electrodes.

The piezoelectric material model is defined as

$$\boldsymbol{s} = \begin{bmatrix} s_{11} & s_{12} & s_{13} & 0 & 0 & 0 \\ s_{21} & s_{22} & s_{23} & 0 & 0 & 0 \\ s_{31} & s_{32} & s_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & s_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & s_{66} \end{bmatrix} \tag{7}$$

$$\boldsymbol{d} = \begin{bmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{24} & 0 & 0 \\ d_{31} & d_{32} & d_{33} & 0 & 0 & 0 \end{bmatrix} \tag{8}$$

$$\boldsymbol{e} = \begin{bmatrix} e_{11} & 0 & 0 \\ 0 & e_{22} & 0 \\ 0 & 0 & e_{33} \end{bmatrix} \tag{9}$$

The output voltage can be calculated from the electric field $\boldsymbol{E_f}$ (Equation (4)) using

$$\boldsymbol{U} = -\int_0^l \boldsymbol{E_f} \cdot d\tilde{l} \tag{10}$$

where l is the distance over which the voltage U is calculated.

According to the literature, the following equalities are applied in Equations (8)–(10), due to polarization and characterization of PZTs (transversely isotropic) [27,37,38]:

$$\begin{aligned} e_{11} &= e_{22} \\ d_{31} &= d_{32} \\ d_{24} &= d_{15} \\ s_{11} &= s_{22} \\ s_{13} &= s_{23} \\ s_{44} &= s_{55} \end{aligned} \tag{11}$$

In our models, the ball is modeled right before the moment of impact with an initial speed calculated by the following free-fall equation:

$$v = \sqrt{2 * g * h} \tag{12}$$

In the above formula, v is the speed at the moment of impact, h is the height that the ball was released from, and g is the gravitational acceleration.

Mass proportional system damping is included in the finite element model, where the acceleration is computed as

$$\boldsymbol{a^n} = \boldsymbol{M}^{-1} \cdot \left(\boldsymbol{P^n} - \boldsymbol{F^n} - \boldsymbol{F^n_{damp}}\right) \tag{13}$$

where $\boldsymbol{M}$ is the diagonal mass matrix, $\boldsymbol{P^n}$ is the external load vector, $\boldsymbol{F^n}$ is the internal load vector, $\boldsymbol{F^n_{damp}}$ is the force vector due to system damping. This latter vector is defined as

$$\boldsymbol{F^n_{damp}} = D_s \cdot \boldsymbol{M} \cdot \boldsymbol{v} \tag{14}$$

where D_s is the system damping constant.

3.1. Fitting Methodology

Since dimensions are directly measured, the input parameters, subject to the fitting procedure of this work, involve the material properties of the plate and PZTs. Tables 2 and 3 present the aforementioned material properties along with their range and step corresponding to a thermoplastic plate and ceramic PZT [41–53]. Range defines the variation between the upper and lower limits of each parameter, while step determines how finely the parameter space is sampled.

Table 2. Range of material properties of PMMA for optimization [41–45].

Property	Symbol	Unit	Range	Step
Young's modulus	E	MPa	1500…65,000	250
Poisson ratio	ν	-	0.3…0.4	0.01

According to Figure 6, the numerical output signal (response) consists of wave propagation, system damping, PZT mechanics, and external excitation (ball to plate impact). Wave propagation is dependent entirely on plate properties (Table 2), while PZT response is based on parameters related to PZT mechanics and system damping (Table 3). Using uncalibrated model parameters will result in a numerical response that significantly deviates from the experimental data collected in Section 2. So, in this work, two fitting subproblems are defined according to Figure 6: (1) plate properties, and (2) PZT properties and system damping. This ensures that initial sampling of uncorrelated properties is not performed, and the subsequent design space of each subproblem is better explored without increasing the computational cost. The aforementioned approach significantly improves the feasibility of this work. While high-performance computing (HPC) was used to accelerate the simulations, the proposed strategy significantly improves feasibility and remains manageable even on standard computational setups.

Table 3. PZT material properties [46–53].

Property	Symbol	Unit	Range	Step
Density	ϱ	$kg/(m^3)$	7500…7800	150
In plane Young's modulus	E_{11}	MPa	55,000…85,000	5000
Out-of-plane Young's modulus	E_{33}	MPa	45,000…75,000	5000
In-plane Poisson ratio	ν_{12}	-	0.3…0.4	0.02
Out-of-plane Poisson ratio	ν_{13}, ν_{23}	-	0.3…0.4	0.02
In plane shear modulus	G_{12}	MPa	20,000…25,000	2500
Out-of-plane shear modulus	G_{13}, G_{23}	MPa	20,000…25,000	2500
PZT coefficient	d_{15}	pC/N	50…800	50
PZT coefficient	d_{31}	pC/N	-800…-50	50
PZT coefficient	d_{33}	pC/N	50…800	50
Relative in-plane permittivity	e_{11}	F/m	500…10,000	250
Relative out-of-plane permittivity	e_{33}	F/m	500…10,000	250
System damping constant	D_s	1/s	200…1200	100

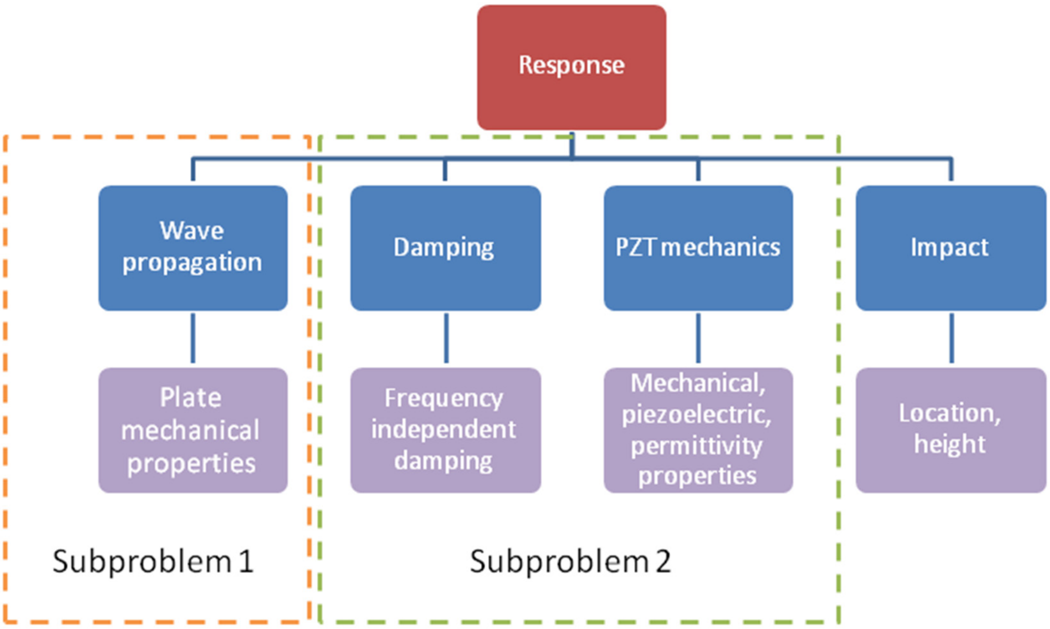


Figure 6. Subproblem definition.

Figure 7 depicts the optimization process used in the current study that utilizes only 4 experimental test cases. Initially, all parameters (Tables 2 and 3) are set at the mean value, while a sequential fitting approach of Subproblem 1 and then Subproblem 2 is performed. Regarding Subproblem 1, full-factorial sampling of the configuration space is performed. Then, by using the FEM, the objective function at all design points is evaluated. The design case exhibiting the lowest error is selected as the optimum solution, according to the grid-search algorithm functionality [54]. Regarding Subproblem 2, the Sobol algorithm is chosen to eliminate subjective bias and allow a good initial sampling of the configuration space (Table 3) [32–34]. Then, by using the FEM, the objective function at all design points is evaluated, and a black box optimization procedure (surrogate optimization) is employed to identify the optimal model parameters. Surrogate optimization is chosen instead of other global optimization algorithms (e.g., simulated annealing, genetic algorithms) as they would result in unreasonable computational cost. The difference in the above approaches lies in the fact that Subproblem 1 exhibits significantly fewer input parameters compared to the number and required sampling of the input parameters in Subproblem 2.

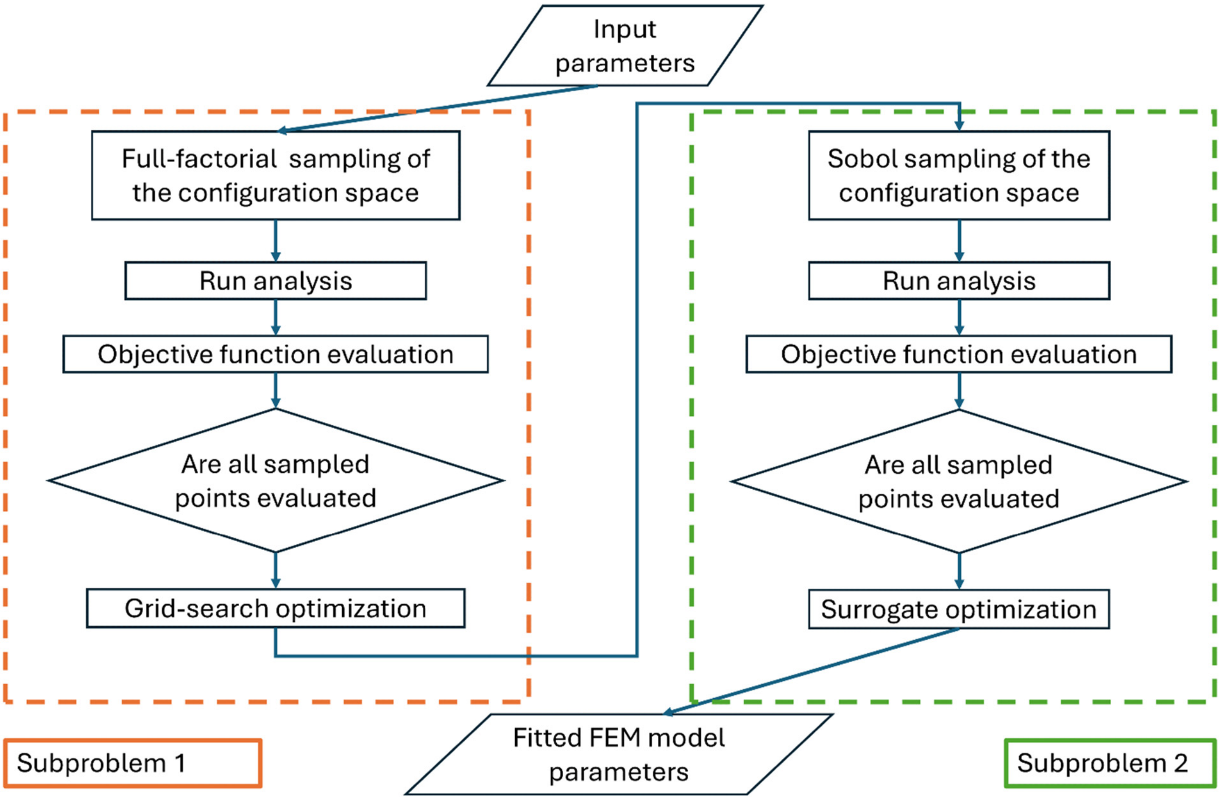


Figure 7. Black box optimization procedure.

Time of flight (ToF) or time of arrival (ToA) is a representative quantity in wave propagation of elastic solids, while many authors include it in impact localization works [20,22,55,56]. In this study, the time of arrival is calculated from the experimental signals for each sensor. Since the actual time of impact is unknown, DToA is used:

$$DToA_{i,j} = ToA_i - ToA_j \text{ and } i \neq j \tag{15}$$

where ToA_i and ToA_j are the times of arrival at sensors i and j, respectively.

The objective function for the plate properties is based on solving nonlinear least squares problems that minimize the objective function $f(p)$, which is a sum of squares:

$$\min_p f(p) = \min_p \sum_{i=1}^{4}\sum_{j=1}^{4}\left(DToA_{i,j}^{exp} - DToA_{i,j}^{sim}\right)^2 \tag{16}$$

where p is the vector of plate parameters to be optimized, $DToA_{i,j}^{exp}$ and $DToA_{i,j}^{sim}$ are the experimental and simulation differences in time of arrival vectors, respectively.

The objective functions for the PZT properties and system damping are based on solving nonlinear least squares problems that minimize the objective functions $f2(p)$ and $f3(p)$, which are a sum of squares:

$$\min_p f2(p) = \min_p \sum_{j=1}^{4}\left(E_j^{exp} - E_j^{sim}\right)^2 \tag{17}$$

where E_j^{exp} and E_j^{sim} are the total experimental and simulation energies of sensor j in the time domain:

$$E = \sum_{n=1}^{n=+\infty}(y_n)^2 \tag{18}$$

$$\min_p f3(p) = \min_p \sum_{j=1}^{4}\left(\frac{Eresid_j^{exp}}{E_j^{exp}} - \frac{Eresid_j^{sim}}{E_j^{sim}}\right)^2 \tag{19}$$

where $Eresid_j^{exp}$ and $Eresid_j^{sim}$ are the experimental and simulation residual energies of sensor j:

$$Eresid = \sum_{n=n1}^{n=n2}(y_n)^2 \tag{20}$$

where y are the sensor values, $n1$ and $n2$ are the first and last indices of the last window of the signal.

The above mathematical formulations and signal processing techniques are graphically depicted in Figure 8. In this, the useful part of the signal (voltage–time) is identified by the use of ToA and the time instance at which the sensor energy converges to a constant value (n2). The obtained signal is segmented into five windows, and the energy of the sensor is calculated (Equation (18)). The residual energy of the sensor (Equation (20)) is calculated in the last window $(n1 < time < n2)$.

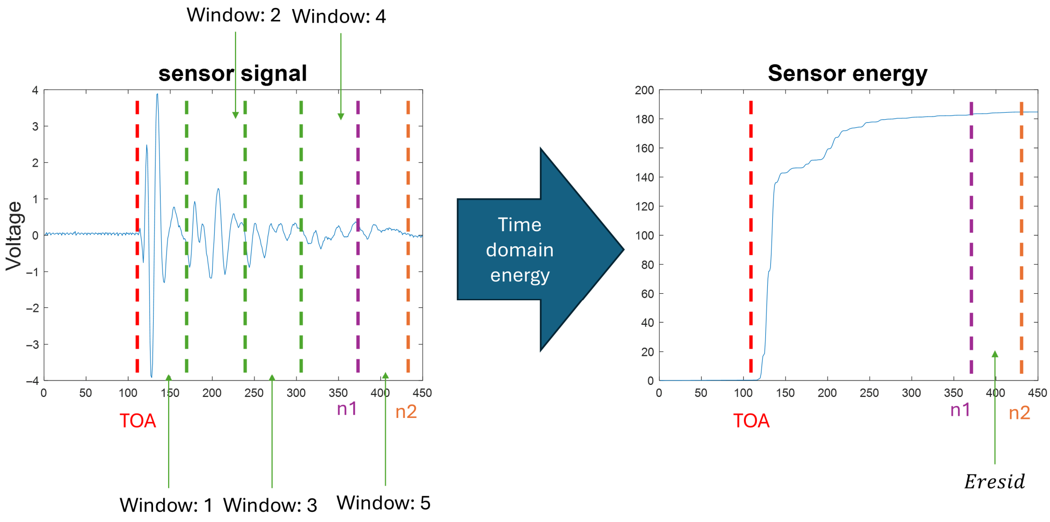


Figure 8. Signal segmentation and energy calculation.

3.2. Data Driven Modeling

The examined surrogate models used in this work involve the following: random forest (RF), Gaussian process regression (GPR), ANN, and singular value decomposition (SVD). The above selection was performed to involve models with low and high interpretability and prediction accuracy, while these models are commonly applied to similar problems [15–17,27,41,57–60].

A random forest regressor averages B regression trees built on bootstrap samples, and at each split, a random subset of features is considered [61,62]. The prediction is the ensemble mean:

$$\hat{f}_{RF}(\boldsymbol{x}) = \frac{1}{B}\sum_{b=1}^{B} T_b(\boldsymbol{x}) \tag{21}$$

where T_b is the b-th tree. Each tree T_b is a regression tree learned by recursive binary partitioning [62]. In accordance with that, let the data reaching the current node be

$$S = \{(\boldsymbol{x_i},\ y_i)\},\ i \in I_s \tag{22}$$

where $\boldsymbol{x_i} = \left(x_{i1}, \ldots, x_{ip}\right)^T \in R^p$ is the predictor vector for observation i and $y_i \in R$ is its response. A candidate split is a feature-threshold pair $s = (j, t)$ with $j \in \{1, \ldots, p\}$ and $t \in R$ that partitions S into two subset 'regions':

$$S_L(j,t) = \left\{i \in I_s :\ x_{ij} \le t\right\}, S_R(j,t) = \left\{i \in I_s :\ x_{ij} > t\right\} \tag{23}$$

As such, a regression tree T_b is grown by choosing a feature j and threshold t that minimize the post-split squared-error impurity:

$$\varepsilon(j,t;S) = \sum_{i\in S_L(j,t)} (y_i - \overline{y}_L)^2 + \sum_{i\in S_R(j,t)} (y_i - \overline{y}_R)^2 \tag{24}$$

where $\overline{y}_L$ and $\overline{y}_R$ are within-region sample means.

The second form of data-driven model of this work involves Gaussian process regression (GPR). According to theory [63], we model an unknown response f with a Gaussian process prior $f \sim GP(m(\cdot), k_\theta(\cdot,\cdot))$. Given inputs $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x_n}]^T \in R^{n\times p}$ and observations $\boldsymbol{Y} = [y_1, \ldots, y_n]^T \in R^n$, the following can be written as

$$y_i = f(\boldsymbol{x_i}) + \varepsilon_i, \qquad \varepsilon_i \sim N\left(0, \sigma_n^2\right) \tag{25}$$

Writing $\boldsymbol{K_\theta} = k_\theta(\boldsymbol{X}, \boldsymbol{X}) \in R^{n\times n}$ for the kernel matrix with

$$\begin{gathered}
[\boldsymbol{K_\theta}]_{ij} = k_\theta(\boldsymbol{x_i}, \boldsymbol{x_j}) \\
\boldsymbol{K_y} = \boldsymbol{K_\theta} + \sigma_n^2\, \boldsymbol{I} \\
\boldsymbol{m}(\boldsymbol{X}) = [m(\boldsymbol{x}_1), \ldots, m(\boldsymbol{x_n})]^T \\
\boldsymbol{k}(\boldsymbol{x}) = [k_\theta(\boldsymbol{x}_1, \boldsymbol{x}), \ldots, k_\theta(\boldsymbol{x_n}, \boldsymbol{x})]^T
\end{gathered} \tag{26}$$

The posterior predictive distribution at a new input $\boldsymbol{x}$ is Gaussian with mean and variance [63]:

$$\mu(\boldsymbol{x}) = \boldsymbol{m}(\boldsymbol{x}) + \boldsymbol{k}(\boldsymbol{x})^T \cdot \boldsymbol{K_y}^{-1} \cdot (\boldsymbol{y} - \boldsymbol{m}(\boldsymbol{X})), \qquad \sigma^2(x) = k_\theta(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}(\boldsymbol{x})^T \cdot \boldsymbol{K_y}^{-1} \cdot \boldsymbol{k}(\boldsymbol{x}) \tag{27}$$

Hyperparameters θ (length-scales, signal variance) are identified by maximizing the log marginal likelihood with gradients available in closed form:

$$\log p(y \mid X, \theta) = -\frac{1}{2}(y - m(X))^\top \cdot K_y{}^{-1} \cdot (y - m(X)) - \frac{1}{2}\log|\boldsymbol{K_y}| - \frac{n}{2}\log(2\pi) \tag{28}$$

The third form of data-driven model is based on classical feedforward neural networks, with Levenberg–Marquardt back propagation chosen as the training algorithm. According to theory [64], the output y_k of the k-th neuron in the output layer is given by

$$y_k = f\left(\sum_{j=1}^{n} w_{kj}^{(2)} \cdot h_j + b_k^{(2)}\right) \tag{29}$$

where $w_{kj}^{(2)}$ is the weight from the j-th neuron in the hidden layer to the k-th neuron in the output layer, h_j is the output of the j-th neuron in the hidden layer, $b_k^{(2)}$ is the bias term for the k-th neuron in the output layer, and f is the activation function used in the output layer.

The output of the j-th neuron in the hidden layer is given by

$$h_j = g\left(\sum_{i=1}^{m} w_{ji}^{(1)} \cdot x_i + b_j^{(1)}\right) \tag{30}$$

where $w_{ji}^{(1)}$ is the weight from the i-th neuron in the input layer to the j-th neuron in the hidden layer, x_i is the input to the i-th neuron in the input layer, $b_j^{(1)}$ is the bias term for the j-th neuron in the hidden layer, and g is the activation function used in the hidden layer.

In this study, a rectified linear unit (ReLU) is used as the activation function, while the loss function is the mean squared error (MSE):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{31}$$

where n is the number of samples or data points, y_i is the actual or observed value for the i-th sample and $\hat{\mathrm{y}}_{\mathrm{i}}$ is the predicted or estimated value for the i-th sample.

Finally, Multivariate Polynomial Interpolation based on the singular value decomposition (SVD) algorithm with variable polynomial degree and normalization is chosen as the last data-driven model [65]. The vector of coefficients $\boldsymbol{C}$ can be obtained by solving the system of linear equations using singular value decomposition:

$$\boldsymbol{A} \cdot \boldsymbol{C} = \boldsymbol{F} \tag{32}$$

where $\boldsymbol{A}$ is the matrix of the input variables, and $\boldsymbol{F}$ is the vector containing the values of the target variable corresponding to each data point.

The SVD of matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \boldsymbol{U} \cdot \boldsymbol{S} \cdot \boldsymbol{V}^T \tag{33}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthogonal matrices, and $\boldsymbol{S}$ is a diagonal matrix containing the singular values. The vector of coefficients $\boldsymbol{C}$ can then be obtained as

$$\boldsymbol{C} = \boldsymbol{V} \cdot \boldsymbol{S}^{-1} \cdot \boldsymbol{U}^T \cdot \boldsymbol{F} \tag{34}$$

In this study, the data are split into 70%, 15% and 15% samples, used for training, validation, and testing, respectively. The Bayesian optimization algorithm is used for hyperparameter estimation, with the search range of each hyperparameter, Table 4, supported by many researchers [15–17,27,41,57–60].

Table 4. Hyperparameter search range of each surrogate model.

Model	Hyperparameter Search Range
RF	Minimum leaf size: [1–150], Number of learners: [5–500]
SVD	Polynomial degree: [2–6], Tolerance: [1×10^{-12}–1×10^{-3}], Standardized data: [true, false]
GPR	Kernel function: [Rational Quadratic, Squared Exponential, Exponential, Mattern 3/2, Mattern 5/2], Standardize data: [true, false]
NN	No. of hidden layers: [1,2,3], Hidden layer size: [1–300], Activation function: [ReLU, Tanh, Sigmoid, None], Standardized data: [true, false]

3.3. Impact Detection

To assess the effectiveness of this study's proposed methodology, an ML approach is utilized to perform impact detection, similar to our previous approach based on experimental data [57]. Specifically, an RF model is utilized to perform regression on the impact location coordinates. This model is an ensemble of regression trees, thus reducing overfitting and effectively handling outliers [66]. In our previous analysis, it was shown to possess the best predictive capabilities out of five different models while it was successfully deployed onto a resource-constrained IoT device, enabling real-time impact monitoring [57].

Regarding impact localization, two different training methodologies are considered in this study. In the first case, training is performed entirely on numerical (FEM) data while testing is performed on experimental data. In the second case, only experimental data are used: the dataset is split into 80% for training and 20% for testing. To ensure a robust methodology, data in the training set, including the four experimental cases used for FEM parameter identification, are not part of the testing set. In addition, the testing set remains consistent regardless of using experimental or simulation data for training. In all cases, the model inputs involve a difference in time of arrival (Equation (15)), as the actual time of impact is unknown. The model outputs involve x, y coordinates. This dual approach allows for evaluating both the representativeness of the synthetic dataset and the model's ability to generalize impact localization across different data domains.

The detection of impact events, based on the principles of wave propagation, follows a specific sensor (i, j) activation sequence. This sequence is determined by the Euclidean distances (s_i, s_j) between the sensors and the point of impact, where sensors located closer to the impact site detect the event before those farther away. As such, to ensure the use of high-quality, noise-free data in our ML framework, impact cases that violate the above are filtered out, and only accepted cases are considered according to the following equation:

$$\begin{gathered} Accepted\ impact\ case: \\ ToA_i \leq ToA_j \text{ if } s_i \leq s_j \quad \forall i, j \in \{1, 2, 3, 4\},\ i \neq j \\ where\ s_i = \sqrt{\left(x_i^{sensor} - X_{impact}\right)^2 + \left(y_i^{sensor} - Y_{impact}\right)^2} \end{gathered} \tag{35}$$

The regression analysis within the context of this work is evaluated by using the coefficient of determination (R^2) and error metrics (root mean square error (RMSE) and mean absolute error (MAE)):

$$\begin{aligned} R^2 &= 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \\ RMSE &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \\ MAE &= \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \end{aligned} \tag{36}$$

where n is the number of data points, y_i is the observed value, $\hat{y}_i$ is the predicted value and $\overline{y}$ is the mean of the observed data. Given that the current study's model predicts two output variables (x, y coordinates), the above metrics are calculated for each output individually and subsequently averaged. This averaging approach ensures a balanced representation of model accuracy across both spatial dimensions, facilitating a concise and meaningful comparison of overall predictive performance [67].

The detection of impact events and the corresponding assessment of impact energy are critical in various engineering applications, where classifying lower and higher velocity impacts is essential for material integrity and structural health monitoring [68]. This is in accordance with the literature, where many researchers have provided corresponding works based on experimental data that utilize ML models to classify and differentiate between low (safe scenario) and high (unsafe scenario) energy impacts [68,69]. As such, in this work, binary classification is employed using a random forest model to distinguish between low and high-impact energy events.

Class creation involves dividing the experimental data based on the 75th percentile of impact energies. In this case, impacts with energy values below the 75th percentile are categorized as low-energy impacts, while those at or above the 75th percentile are classified as high-energy impacts. This method ensures that the classification is data-driven and adaptable to the distribution of the impact energies, providing a balanced and representative separation of classes. Finally, the energy recorded by each sensor serves as input to the ML model, while the output corresponds to the assigned impact energy class, enabling accurate differentiation between low and high energy events based on sensor data.

Similarly to our regression approach, training is performed entirely on numerical (FEM) data while testing is performed on experimental data. To test the effectiveness of this approach, error metrics are provided based on the use of only experimental data for both training and testing. In the last case, the experimental data are split into 80–20%, used for training and testing, respectively. Classification performance is evaluated based on accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{37}$$

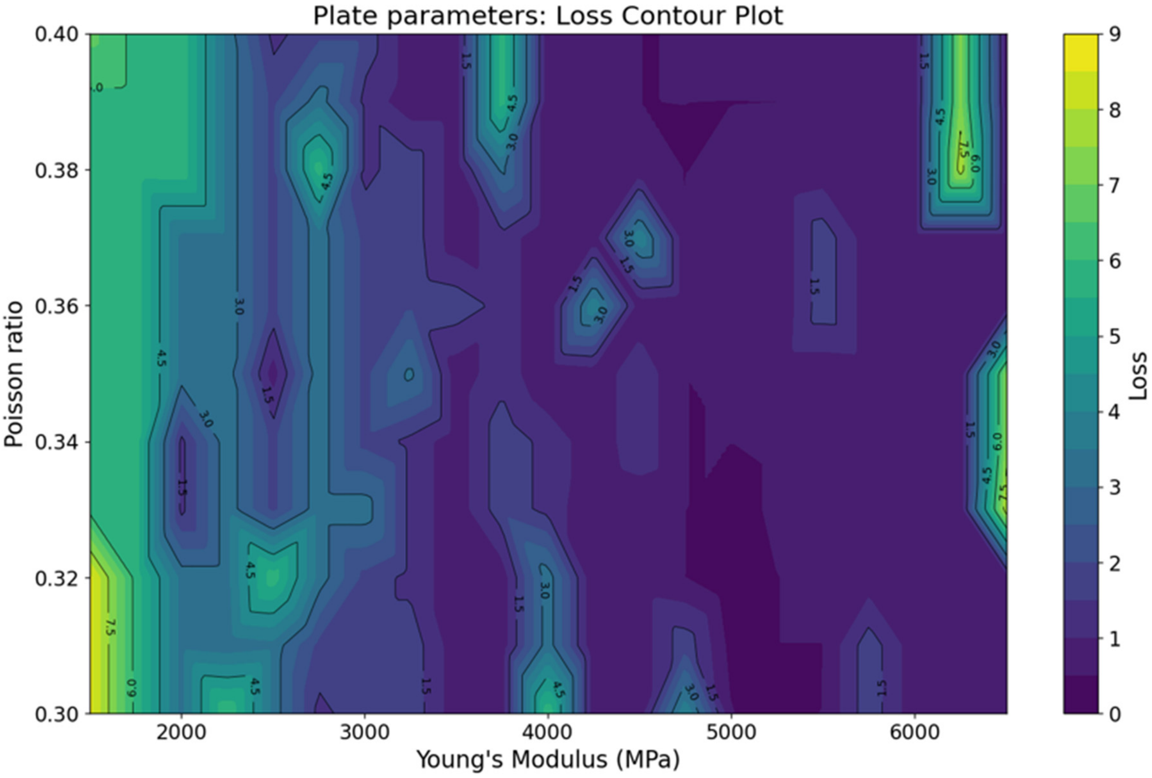
where TP are the True Positives (correctly predicted positive cases), TN are the True Negatives (correctly predicted negative cases), FP are the False Positives (incorrectly predicted positive cases), FN are the False Negatives (incorrectly predicted negative cases).

4. Results

4.1. Fitting Results

Figure 9 presents the results of the fitting procedure used to estimate plate properties by utilizing full factorial sampling. The loss function values correspond to differences in DToA indices between experimental and simulation data, as presented by Equation (15). The contour plot represents the iso-loss surface as a function of Young's modulus and

Poisson ratio, allowing visualization of the sensitivity of the calibration process to the material parameters. The design case exhibiting the minimum loss with the corresponding optimal plate properties is given in the same Figure.



Full factorial sampling: Optimal plate properties	
Young's modulus (MPa)	5000
Poisson ratio	0.32

Figure 9. Full factorial sampling results for plate parameters fitting.

As presented in Figure 7, the estimation of PZT properties and system damping is performed by a significantly more complex optimization methodology compared to plate properties identification. As such, it is essential to verify the quality of the initial sampling, as this directly influences the reliability of the surrogate models. The correlations among the input parameters of Subproblem 2 were examined and found to be negligible, with all absolute correlation coefficients below 0.07. This indicates that the Sobol sampling procedure effectively produced an uncorrelated input space, ensuring that each variable contributes independently to the exploration of the design space. Such independence is crucial for avoiding biased surrogate model training and for maintaining the validity of subsequent parameter identification results.

To examine the effect of each parameter on each output (explained in Section 3), a factor analysis is performed using the Student method [70]. Input influence is screened using a two-group contrast and Student's *t*-test. For each factor, the runs are split into a 'low' and a 'high' set. The effect is the signed difference in the two mean responses, and a two-sample *t*-test (Welch) assesses whether that difference is non-zero. Effects are reported in the response's own units (positive = direct, negative = inverse). This is the main effect definition for two-level factorial designs; two-factor interactions are obtained analogously. Compared with Pearson/Spearman correlations (unitless linear/monotone association on all data) and Sobol's indices (global variance fractions without sign and high sample cost), the t-based approach is a lightweight screen that yields direction and statistical significance [70–73].

Figure 10 shows the main factors (high significance) of each output variable (left: sensor energies, right: residual sensor energies). Sensor energies are found to be mostly dependent on the piezoelectric properties affecting the output voltage in the z-direction, which is in accordance with the poling direction of the PZTs. Higher damping translates to more dispersed energy during wave propagation. As such, this parameter significantly affects total sensor energy, while dominating residual sensor energy (the effect of all other parameters is negligible). The above are in accordance with laws of piezoelectricity, wave propagation, and further validate the numerical modeling and choice of objective functions used to describe and simplify the numerical framework of this work.

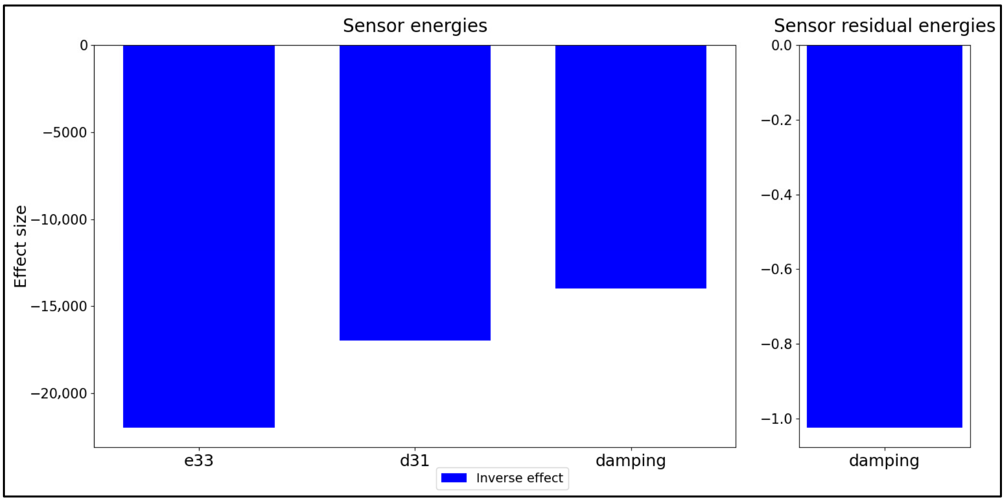


Figure 10. Factor analysis (Fitting Subproblem 2).

Table 5 indicates the predictive or testing set performance of each RSM on each output. In terms of residual sensor energies, the SVD outperforms all other models, while both the NN and the Kriging models show low accuracy results. Regarding sensor energies, the NN provides excellent results in contrast to SVD and Kriging models, which are deemed unacceptable.

Table 5. Predictive performance of each RSM.

Output Variable	R^2 Kriging	R^2 NN	R^2 SVD
Residual sensor energies	0.56	0.34	0.82
Sensor energies	0.48	0.97	0.54

Using the above data-driven models, the optimal FEM parameters are identified and presented in Table 6. Specifically, given the dominance of system damping on the residual sensor energies, the damping constant is identified by using the SVD model (output variable: Residual sensor energies). Then, using the NN model (output variable: sensor energies), the rest of the parameters are identified.

Table 6. Material parameters of PZT determined by optimization (Fitting Subproblem 2).

Parameter	Unit	Fitted Value
E_{11}	MPa	75,200
E_{33}	MPa	55,200
G_{12}	MPa	24,800
G_{13}	MPa	24,900
ν_{12}	-	0.3
ν_{13}	-	0.34
ρ	kg/(m^3)	7531.4
D_s	1/s	1086
e_{11}	F/m	3875.5
e_{33}	F/m	6320.2
d_{15}	pC/N	50
d_{31}	pC/N	−737.8
d_{33}	pC/N	793

4.2. Impact Identification Results

Tables 7 and 8 and Figure 11 illustrate the regression model's testing set performance in localizing impacts across two different regions: the entire plate and the inner region (excluding edge impacts). When using simulated data for training and experimental data for testing, throughout the whole plate (Table 7), the model achieves an R^2 of 0.81, a RMSE of 30.4 mm, and an MAE of 22.4 mm on the test data. These results are comparable to those obtained when both training and testing use experimental data, indicating that training with simulated data yields similar localization accuracy. The performance further improves when focusing on the interior region (Table 8), with an R^2 of 0.89, RMSE of 20.9 mm, and MAE of 17 mm, again showing that both simulated and experimental training data provide similar outcomes. Figure 11 confirms that predicted values (in red) closely follow the actual values (in blue) in all cases (impact case number given as an integer next to points), with the error decreasing when edge impacts are excluded.

Table 7. Localization results for impact cases in all of the plates.

Case	R^2	RMSE (mm)	MAE (mm)
Training: simulation Testing: experimental	0.81	30.4	22.4
Training: experimental Testing: experimental	0.82	30	22.4

Table 8. Localization results for the inside region (excluding edge cases).

Case	R^2	RMSE (mm)	MAE (mm)
Training: simulation Testing: experimental	0.9	20.1	15.5
Training: experimental Testing: experimental	0.9	18.5	14.9

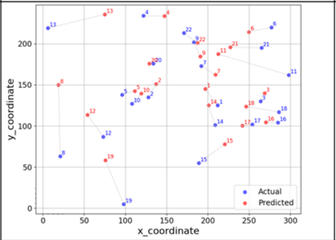
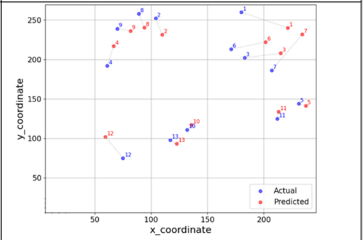
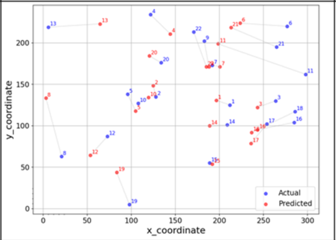
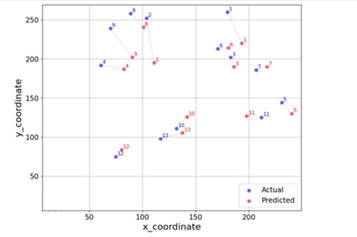
Case	All plate	Inside region
Training: Simulation **Testing: Experimental**		
Training: Experimental **Testing: Experimental**		

Figure 11. Predicted vs. true plots across training–testing dataset combinations. Integers next to points denote impact case number.

Figure 12 presents a comparison of average localization errors (mean absolute error in mm) across several studies, including our own. The results of our work, based on simulation data, align well with the findings of other researchers who utilize experimental data. As observed, testing not involving impact cases near the edges (inside region) consistently yields lower localization errors compared to testing across the whole plate (all plate), further validating the methodology of this work [16,18,21,74–77].

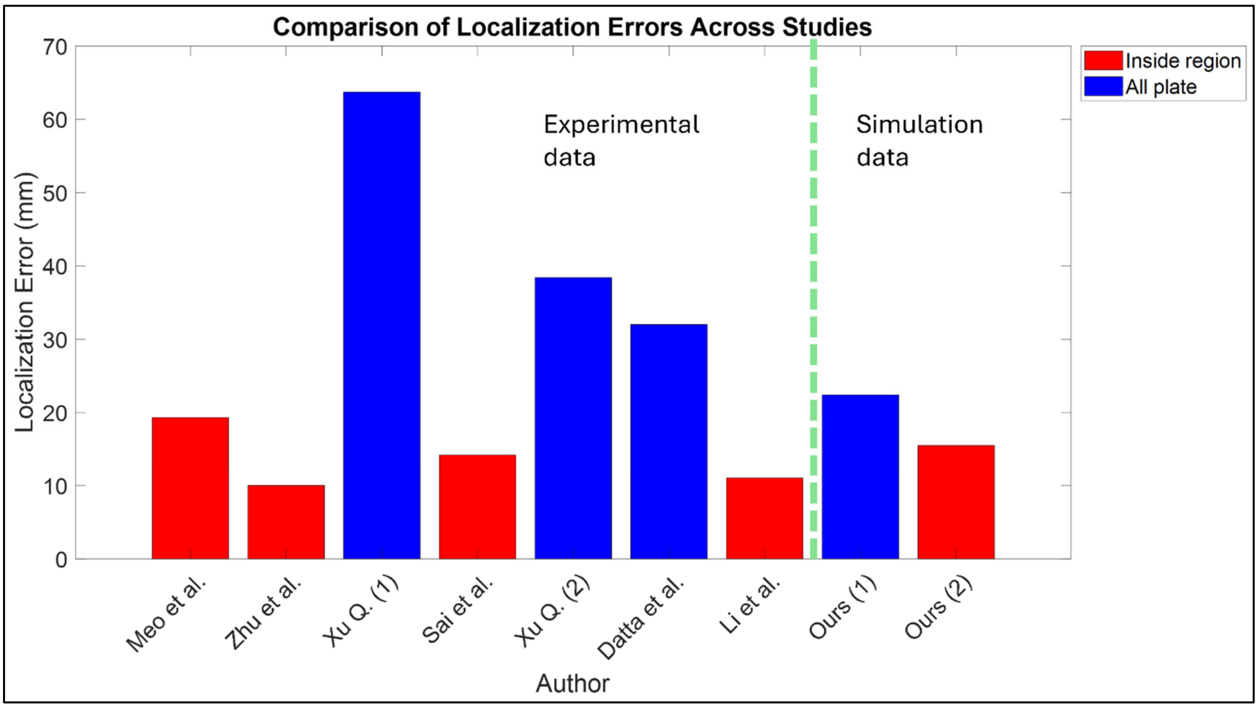


Figure 12. Comparative literature localization errors [16–18,21,74–76].

Table 9 presents the classification performance of the current study. Notably, the results show minimal variation between the examined regions and the use of the training dataset. Specifically, an accuracy of 0.78 is achieved when using simulation data compared to the accuracies of the experimental training set (inner region: 0.78, all plate: 0.83). The above indicates that the numerical framework of this work effectively captures the key features necessary for classification, underscoring its robustness in providing generalizable insights for impact classification tasks.

Table 9. Impact classification results across training–testing dataset combinations.

Case	Accuracy	
	All Plate	**Inside Region**
Training: simulation Testing: experimental	0.78	0.78
Training: experimental Testing: experimental	0.78	0.83

5. Conclusions

In this study, an end-to-end numerical methodology for accurate impact detection based on simulation (training) and experimental (testing) data is presented. Specifically, an impact detection setup utilizing PZTs as sensors mounted on a thermoplastic plate is used to obtain experimental signals at different impact locations and velocities in lab conditions. A corresponding FEM is created, and its parameters (material and piezoelectric properties) are identified using grid search and surrogate optimization. Then, a virtual dataset, involving multiple impact cases, is created and used to train ML models performing impact detection. The results indicate the following.

Plate material properties ($E = 5000$ MPa, $\nu = 0.32$) are identified by full factorial sampling and minimizing the differences in DToA indices between experimental and simulation data.

Regarding PZT properties and system damping, the Student method indicates high significance factors for each output variable (sensor energies: e_{33}, d_{31}, and $damping$) (residual sensor energies: $damping$).

PZT properties and system damping are identified using RSM, in which the SVD outperforms all other models on residual sensor energies, while the NN provides excellent results on sensor energies.

The regression model's performance in localizing impacts is similar when trained with either experimental or simulation data, while performance is increased for both cases in the inner region of the plate. The presented localization performance aligns well with the findings of other researchers utilizing experimental data.

Classification between low and high velocity impacts shows acceptable accuracy and minimal variation between the examined regions and the use of the training dataset.

This study aims to provide a systematic, repeatable numerical framework for event representative dataset generation and impact monitoring through ML. It also addresses accurate FEM parameter identification from a few experimental tries, without extensive testing of all system components. The developed datasets and ML models are not expected to perform accurately in other setups, given that monitoring systems differ in terms of sensor selection and placement, monitored structure, and applied loads. However, the presented methodology can assist by providing a standardized approach to data preprocessing, feature extraction, and model training that can be adapted to other setups with minimal modifications.

Author Contributions: Conceptualization, P.G., J.G., C.T., S.H.P., A.K., I.K. and A.H.; methodology, P.G., J.G., C.T., S.H.P., S.G., V.K. and S.N.; software, P.G., C.T. and S.H.P.; validation, P.G., J.G., C.T. and S.H.P.; formal analysis, P.G. and J.G.; investigation, J.G., C.T. and S.H.P.; resources, S.G., V.K. and S.N.; data curation, P.G.; writing—original draft preparation, P.G.; writing—review and editing, J.G. and M.G.; visualization, P.G.; supervision, A.K., I.K., A.H., M.G., V.K. and S.N.; project administration,

A.K., I.K. and M.G.; funding acquisition, A.K., I.K., A.H. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by COST action HISTRATE under grant agreement CA21155, the Andreas Mentzelopoulos foundation and by the cooperation and education project M2BRIDGE–Towards a Sustainable Cooperation in Teaching and Research in Lightweight Engineering and Multi-Material Design between TU Dresden and the University of Patras, financed by the German Academic Exchange Service (DAAD) with funds from the Federal Foreign Office (Auswärtiges Amt, AA), in the framework of the German-Greek Academic Partnerships 2023–2025.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw/processed data required to reproduce these findings will be made available upon reasonable request.

Conflicts of Interest: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	artificial neural network.
SVM	support vector machine.
DOE	design of experiment.
DToA	difference in time of arrival.
FEM	finite element model.
GPR	Gaussian process regression.
MAE	mean absolute error.
ML	machine learning.
ToA	time of arrival.
ToF	time of flight.
PMMA	polymethyl methacrylate.
PZT	piezoelectric transducer.
R^2	coefficient of determination
RF	random forest.
RMSE	root mean square error.
SHM	structural health monitoring.
SVD	singular value decomposition
RSM	response surface methodology.

References

1. Azimi, M.; Eslamlou, A.; Pekcan, G. Data-driven structural health monitoring and damage detection through deep learning: State-of-the art review. *Sensors* **2020**, *20*, 2778. [CrossRef]
2. Sony, S.; Laventure, S.; Sadhu, A. A literature review of next generation smart sensing technology in structural health monitoring. *Struct. Control Health Monit.* **2019**, *26*, e2321. [CrossRef]
3. Monti, G.; Rabi, R.R.; Marella, L.; Proietti, S.T. Data-driven decision support system for the safety management of railway bridge networks. *Reliab. Eng. Syst. Saf.* **2025**, *262*, 111202. [CrossRef]
4. Han, H.; Meng, Z.; Shi, B.; Zha, F.; Wei, G.; Yu, L.; Song, X. A method for monitoring of monopile horizontal displacement of offshore wind turbine based on UWFBG and boundary reconstruction. *Measurement* **2025**, *252*, 117384. [CrossRef]
5. Gui, G.; Pan, H.; Lin, Z.; Li, Y.; Yuan, Z. Data-driven support vector machine with optimization techniques for structural health monitoring and damage detection. *Ksce J. Civ. Eng.* **2017**, *21*, 523–534. [CrossRef]
6. Tselios, I.; Nikolakopoulos, P. Combining artificial neural networks and mathematical models for unbalance estimation in a rotating system under the nonlinear journal bearing approach. *Lubricants* **2024**, *12*, 344. [CrossRef]

7. Varouxis, T.; Tserpes, K.; Fassois, S. Progressive fatigue damage detection and assessment in composite specimens using random vibration response signals. In Proceedings of the 20th European Conference on Composite Materials (ECCM20), Lausanne, Switzerland, 26–30 June 2022. [CrossRef]
8. Andreades, C.; Fierro, G.P.M.; Meo, M. A nonlinear ultrasonic shm method for impact damage localisation in composite panels using a sparse array of piezoelectric pzt transducers. *Ultrasonics* **2020**, *108*, 106181. [CrossRef] [PubMed]
9. Yue, N.; Khodaei, Z.S.; Aliabadi, M. Damage detection in large composite stiffened panels based on a novel shm building block philosophy. *Smart Mater. Struct.* **2021**, *30*, 045004. [CrossRef]
10. Capineri, L.; Bulletti, A. A versatile analog electronic interface for piezoelectric sensors used for impacts detection and positioning in structural health monitoring (shm) systems. *Electronics* **2021**, *10*, 1047. [CrossRef]
11. Capineri, L.; Bulletti, A. Ultrasonic guided-waves sensors and integrated structural health monitoring systems for impact detection and localization: A review. *Sensors* **2021**, *21*, 2929. [CrossRef]
12. Guo, Z.; Huang, T.; Schröder, K.-U. Development of a piezoelectric transducer-based integrated structural health monitoring system for impact monitoring and impedance measurement. *Appl. Sci.* **2020**, *10*, 2062. [CrossRef]
13. Dipietrangelo, F.; Nicassio, F.; Scarselli, G. Shm implementation on a rpv airplane model based on machine learning for impact detection. *Aerotec. Missili Spaz.* **2024**, *103*, 363–375. [CrossRef]
14. Liu, Q.; Wang, F.; Li, J.; Xiao, W. A hybrid support vector regression with multi-domain features for low-velocity impact localization on composite plate structure. *Mech. Syst. Signal Process.* **2021**, *154*, 107547. [CrossRef]
15. Sai, Y.; Zhao, X.; Wang, L.; Hou, D. Impact localization of cfrp structure based on fbg sensor network. *Photonic Sens.* **2019**, *10*, 88–96. [CrossRef]
16. Xu, Q. A comparison study of extreme learning machine and least squares support vector machine for structural impact localization. *Math. Probl. Eng.* **2014**, *2014*, 906732. [CrossRef]
17. Datta, A.; Augustin, M.J.; Gupta, N.; Viswamurthy, S.R.; Gaddikeri, K.M.; Sundaram, R. Impact localization and severity estimation on composite structure using fiber bragg grating sensors by least square support vector regression. *IEEE Sens. J.* **2019**, *19*, 4463–4470. [CrossRef]
18. Jang, B.-W.; Kim, C.-G. Acoustic emission source localization in composite stiffened plate using triangulation method with signal magnitudes and arrival times. *Adv. Compos. Mater.* **2020**, *30*, 149–163. [CrossRef]
19. Shrestha, P.; Kim, J.-H.; Park, Y.; Kim, C.-G. Impact localization on composite structure using fbg sensors and novel impact localization technique based on error outliers. *Compos. Struct.* **2016**, *142*, 263–271. [CrossRef]
20. Li, H.; Wang, Z.; Forrest, J.Y.-L.; Jiang, W. Low-velocity impact localization on composites under sensor damage by interpolation reference database and fuzzy evidence theory. *IEEE Access* **2018**, *6*, 31157–31168. [CrossRef]
21. Jang, B.-W.; Kim, C.-G. Impact localization of composite stiffened panel with triangulation method using normalized magnitudes of fiber optic sensor signals. *Compos. Struct.* **2019**, *211*, 522–529. [CrossRef]
22. Jang, B.-W.; Lee, Y.-G.; Kim, C.-G.; Park, C.-Y. Impact source localization for composite structures under external dynamic loading condition. *Adv. Compos. Mater.* **2015**, *24*, 359–374. [CrossRef]
23. Miele, S.; Karve, P.; Mahadevan, S. Multi-fidelity physics-informed machine learning for probabilistic damage diagnosis. *Reliab. Eng. Syst. Saf.* **2023**, *235*, 109243. [CrossRef]
24. Li, M.; Wu, Z.; Yang, H.; Huang, H. Direct damage index based on inverse finite element method for structural damage identification. *Ocean Eng.* **2021**, *221*, 108545. [CrossRef]
25. Ye, Z.; Hsu, S.-C. Predicting real-time deformation of structure in fire using machine learning with cfd and fem. *Autom. Constr.* **2022**, *143*, 104574. [CrossRef]
26. Mirasoli, G.; Brutti, C.; Groth, C.; Mancini, L.; Porziani, S.; Biancolini, M.E. Structural health monitoring of civil structures through fem high-fidelity modelling. *IOP Conf. Ser. Mater. Sci. Eng.* **2022**, *1214*, 012019. [CrossRef]
27. DHesser, F.; Kocur, G.K.; Markert, B. Active source localization in wave guides based on machine learning. *Ultrasonics* **2020**, *106*, 106144. [CrossRef]
28. Ren, L.; Zhong, Y.; Xiang, J.; Wang, Z. Adaptive sensor array error calibration based impact localization on composite structure. *Appl. Sci.* **2020**, *10*, 4042. [CrossRef]
29. Katsidimas, I.; Kotzakolios, T.; Nikoletseas, S.; Panagiotou, S.H.; Timpilis, K.; Tsakonas, C. Impact events for structural health monitoring of a plastic thin plate: Dataset. In Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems, SenSys'22, Boston, MA, USA, 6–9 November 2022; Association for Computing Machinery: New York, NY, USA, 2023; pp. 1020–1025.
30. Experimental Impact Events Dataset. Available online: https://zenodo.org/record/7199346 (accessed on 7 March 2025).
31. CUI Devices. CEB-35D26 Datasheet. Available online: https://gr.mouser.com/datasheet/2/670/ceb_35d26-1776373.pdf (accessed on 7 March 2025).
32. Gkertzos, P.; Kotzakolios, A.; Katsidimas, I.; Kostopoulos, V. Parametric numerical study and multi-objective optimization of composite curing through infrared radiation. *Appl. Mech.* **2024**, *5*, 192–211. [CrossRef]

33. Sobol', I.M.; Asotsky, D.; Kreinin, A.; Kucherenko, S. Construction and comparison of high-dimensional sobol' generators. *Wilmott* **2011**, *2011*, 64–79. [CrossRef]
34. Bratley, P.; Fox, B.L. Algorithm 659: Implementing sobol's quasirandom sequence generator. *ACM Trans. Math. Softw.* **1988**, *14*, 88–100. [CrossRef]
35. LS-DYNA Official Site. Available online: https://www.dynasupport.com (accessed on 7 March 2025).
36. Nirbhay, M.; Dixit, A.; Misra, R. Finite element modelling of lamb waves propagation in 2d plates and thin sheets for damage detection. *Mater. Werkst* **2017**, *48*, 577–588. [CrossRef]
37. Zhou, S.; Zhang, R.; Li, A.; Qiao, J.; Zhou, S. Analysis of transversely isotropic piezoelectric bilayered rectangular micro-plate based on couple stress piezoelectric theory. *Eur. J. Mech.-A/Solids* **2022**, *96*, 104707. [CrossRef]
38. Liu, M.; Yang, F. Finite element analysis of the spherical indentation of transversely isotropic piezoelectric materials. *Model. Simul. Mater. Sci. Eng.* **2012**, *20*, 045019. [CrossRef]
39. MStengel; Spaldin, N.A.; Vanderbilt, D. Electric displacement as the fundamental variable in electronic-structure calculations. *Nat. Phys.* **2009**, *5*, 304–308. [CrossRef]
40. Li, J.F. Fundamentals of Piezoelectricity. In *Lead-Free Piezoelectric Materials*; Wiley-VCH GmbH: Weinheim, Germany, 2021; pp. 1–18.
41. Gkertzos, P.; Kotzakolios, A.; Mantzouranis, G.; Kostopoulos, V. Nozzle temperature calibration in 3d printing. *IJIDEM* **2023**, *18*, 879–899. [CrossRef]
42. Modulus of Elasticity and Poisson's Coefficient of Polymeric Materials. Available online: https://www.sonelastic.com/en/fundamentals/tables-of-materials-properties/polymers.html (accessed on 7 March 2025).
43. Thermoplastics-Physical Properties. Available online: https://www.engineeringtoolbox.com/physical-properties-thermoplastics-d_808.html (accessed on 7 March 2025).
44. Grigore, M. Methods of recycling, properties and applications of recycled thermoplastic polymers. *Recycling* **2017**, *2*, 24. [CrossRef]
45. Ibeh, C.C. *Thermoplastic Materials*; CRC Press: Boca Raton, FL, USA, 2011.
46. Piezoceramic Materials. Available online: https://www.piceramic.com/en/expertise/piezo-technology/piezoelectric-materials#c15065 (accessed on 7 March 2025).
47. Omega Piezo Manufactured Piezoelectric Ceramic Components. Available online: https://www.omegapiezo.com/ceramic-components/ (accessed on 7 March 2025).
48. Material: Lead Zirconate Titanate (PZT). Available online: https://www.memsnet.org/material/leadzirconatetitanatepzt/ (accessed on 7 March 2025).
49. APC International: Physical & Piezoelectric Properties. Available online: https://www.americanpiezo.com/apc-materials/piezoelectric-properties.html (accessed on 7 March 2025).
50. PI Ceramic: Material Data. Available online: https://www.piceramic.com/fileadmin/user_upload/physik_instrumente/files/datasheets/PI_Ceramic_Material_Data.pdf (accessed on 7 March 2025).
51. Carrino, S.; Nicassio, F.; Scarselli, G. Nonlinearities associated with impaired sensors in a typical shm experimental set-up. *Electronics* **2018**, *7*, 303. [CrossRef]
52. Krommer, M.; Berik, P.; Vetyukov, Y.; Benjeddou, A. Piezoelectric d15 shear-response-based torsion actuation mechanism: An exact 3d saint-venant type solution. *Int. J. Smart Nano Mater.* **2012**, *3*, 82–102. [CrossRef]
53. Poisson Ratio for Poled Electroceramics. Available online: https://apps.dtic.mil/sti/tr/pdf/ADA299045.pdf (accessed on 7 March 2025).
54. Understanding Grid Search as an Optimization Algorithm in Machine Learning. Available online: https://neurosnap.ai/blog/post/643748be49872f3862f39aed (accessed on 7 March 2025).
55. Achenbach, J.D. *Wave Propagation in Elastic Solids*; North-Holland: Amsterdam, The Netherlands, 1973.
56. Ciampa, F.; Meo, M.; Barbieri, E. Impact localization in composite structures of arbitrary cross section. *Struct. Health Monit.-Int. J.* **2012**, *11*, 643–655. [CrossRef]
57. Katsidimas, I.; Kostopoulos, V.; Kotzakolios, T.; Nikoletseas, S.; Panagiotou, S.; Tsakonas, C. An impact localization solution using embedded intelligence—Methodology and experimental verification via a resource-constrained iot device. *Sensors* **2023**, *23*, 896. [CrossRef]
58. Ullmann, F.; Hardt, W.; Zhmud, V. Machine learning algorithms for impact localization on formed piezo metal composites. In Proceedings of the 2017 International Siberian Conference on Control and Communications (SIBCON), Astana, Kazakhstan, 29–30 June 2017; pp. 1–5.
59. Gkertzos, P.; Kotzakolios, A.; Mantzouranis, G.; Kostopoulos, V. Effect of slicing parameters on the as-manufactured state of 3d printed parts utilizing numerical modeling. *Int. J. Adv. Manuf. Technol.* **2024**, *135*, 4879–4909. [CrossRef]
60. Gkertzos, P.; Kotzakolios, A.; Kostopoulos, V. Multi-parametric numerical analysis of 3d printed sparse infill structures. *Int. J. Adv. Manuf. Technol.* **2024**, *134*, 1143–1167. [CrossRef]
61. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

62. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C.J. *Classification and Regression Trees*; Chapman and Hall/CRC: New York, NY, USA, 1984.
63. Rasmussen, C.E.; Williams, C.K.I. *Gaussian Processes for Machine Learning*; The MIT Press: Cambridge, MA, USA, 2005.
64. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
65. Golub, G.H.; Van Loan, C.F. *Matrix Computations*, 4th ed.; Johns Hopkins University Press: Philadelphia, PA, USA, 2013.
66. Advantages and Disadvantages of Random Forest. Available online: https://www.geeksforgeeks.org/what-are-the-advantages-and-disadvantages-of-random-forest/ (accessed on 7 March 2025).
67. Dugalam, R.; Prakash, G. A hybrid multiple input multiple output (mimo) model for simultaneous localization and quantification of structural damage in beam structures. *Structures* **2024**, *60*, 105879. [CrossRef]
68. Nucera, C.; White, S.; Chen, Z.M.; Kim, H.; Lanza, F. Impact monitoring in stiffened composite aerospace panels by wave propagation. *Struct. Health Monit.* **2015**, *14*, 547–557. [CrossRef]
69. Damm, A.M.; Spitzmüller, C.; Raichle, A.T.S.; Bühler, A.; Weißgraeber, P.; Middendorf, P. Deep learning for impact detection in composite plates with sparsely integrated sensors. *Smart Mater. Struct.* **2020**, *29*, 125014. [CrossRef]
70. Student. The probable error of a mean. *Biometrika* **1908**, *6*, 1–25. [CrossRef]
71. Sedgwick, P. Pearson's correlation coefficient. *BMJ* **2012**, *345*, e4483. [CrossRef]
72. Wissler, C. The spearman correlation formula. *Science* **1905**, *22*, 309–311. [CrossRef]
73. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Sensitivity Analysis: From Theory to Practice, ch. 6*; John Wiley Sons, Ltd.: Hoboken, NJ, USA, 2007; pp. 237–275.
74. Meo, M.; Zumpano, G.; Piggott, M.; Marengo, G. Impact identification on a sandwich plate from wave propagation responses. *Compos. Struct.* **2005**, *71*, 302–306. [CrossRef]
75. Zhu, K.; Qing, X.P.; Liu, B. A two-step impact localization method for composite structures with a parameterized laminate model. *Compos. Struct.* **2018**, *192*, 500–506. [CrossRef]
76. Xu, Q. Impact detection and location for a plate structure using least squares support vector machines. *Struct. Health Monit. Int. J.* **2013**, *13*, 5–18. [CrossRef]
77. Abdi, M.; Sorokin, V.; Mace, B. Reflection of waves in a waveguide from a boundary with nonlinear stiffness: Application to axial and flexural vibrations. *Nonlinear Dyn.* **2022**, *109*, 3051–3082. [CrossRef]