ACCSCIENCE PUBLISHING

International Journal of AI for Materials and Design

ORIGINAL RESEARCH ARTICLE

Physics-guided machine learning for ultrasonic crack-depth quantification in laboratory rail specimens using laser ultrasonics

Faeez Masurkar*

The British University in Dubai, Dubai, United Arab Emirates

Abstract

Accurate quantification of crack depth is essential for ensuring the structural integrity and operational safety of railway infrastructure. Conventional ultrasonic inspection techniques primarily focus on defect detection and often require baseline measurements or expert interpretation, limiting their suitability for autonomous structural health monitoring. This study proposes a physics-guided machine learning (ML) framework for baseline-free ultrasonic crack-depth quantification in rail-track structures using ultrasonic guided-wave and data-driven regression models. Rayleigh-wave ultrasonic signals were acquired from 13 rail specimens, comprising three healthy specimens and ten specimens with crack depths ranging from 2 to 20 mm. Each specimen was tested 25 times, resulting in a database of 325 ultrasonic measurements. After signal preprocessing, nine physics-guided damage-sensitive features, including amplitude, energy, attenuation, arrival time, and wavelet-based descriptors, were extracted to characterize crack-induced changes in wave propagation. Four regression algorithms, namely Gradient Boosted Regression Trees (GBRT-LSBoost), Random Forest (RF), Support Vector Regression (SVR), and Gaussian Process Regression (GPR), were evaluated using independent test data and specimen-wise, grouped cross-validation. Model interpretability and robustness were further assessed through feature importance, permutation importance, feature ablation, correlation analysis, and uncertainty quantification. The results show that amplitude and attenuation-based features are the most influential predictors of crack depth, whereas time-of-flight and several wavelet features contribute marginally. Among the four nonlinear ML models, Random Forest achieved the best independent-test performance (R^2 = 0.961, RMSE = 1.610 mm), closely followed by GBRT-LSBoost (R^2 = 0.960, RMSE = 1.623 mm) which could be suitable for field deployment under complex and dynamic conditions. Notably, a peak-amplitude-only linear regression baseline achieved lower error (R^2 = 0.995, RMSE = 0.563 mm), demonstrating the strong approximately linear relationship between transmitted Rayleigh-wave amplitude and prescribed crack depth under the controlled laboratory conditions. The results demonstrate the feasibility of the proposed feature-level physics-guided regression framework under controlled laboratory conditions, while further validation using a larger number of independent specimens and naturally occurring service cracks is required before field deployment.

Keywords: Physics-guided machine learning; Ultrasonic guided waves; Rail-track structural health monitoring; Crack-depth quantification; Rayleigh wave ultrasonics; Non-destructive testing

***Corresponding author:**
Faeez Masurkar
(faeez.masurkar@buid.ac.ae)

Citation: Masurkar F. Physics-guided machine learning for ultrasonic crack-depth quantification in laboratory rail specimens using laser ultrasonics. *Int J AI Mater Design*. 2026;3(3):026310028. doi: 10.36922/IJAMD026310028

Received: July 30, 2026

Revised: September 4, 2026

Accepted: September 11, 2026

Published online: September 17, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Railway transportation is one of the most efficient and economical modes of passenger and freight transport. However, increasing axle loads, train speeds, and traffic density have imposed greater demands on the safety and reliability of railway infrastructure. During service, rails are subjected to cyclic wheel–rail contact forces, thermal loading, residual manufacturing stresses, and environmental degradation, leading to the initiation and propagation of cracks. Common defects include transverse cracks, head checks, squats, bolt-hole cracks, weld defects, rolling contact fatigue, and internal fissures. If left undetected, these defects can propagate and ultimately cause rail fracture and catastrophic derailment. Consequently, early crack detection and quantitative damage assessment have become major research priorities in railway structural health monitoring and nondestructive testing.[1–4]

Conventional rail inspection methods, including visual inspection, magnetic particle testing, eddy current testing, radiography, and ultrasonic pulse–echo testing, are widely used for defect detection. However, visual, magnetic particles, and eddy current techniques are limited to surface or near-surface defects, whereas radiography is expensive and involves radiation hazards. Although ultrasonic pulse–echo testing remains the industry standard for internal defect detection, its point-by-point scanning approach is time-consuming and its performance can be affected by rail joints, welds, and geometric discontinuities.[1,5]

To overcome these limitations, ultrasonic guided waves have emerged as an effective technique for long-range rail track inspection. Unlike bulk waves, guided waves propagate along the rail while remaining confined within their boundaries, allowing several meters of track to be inspected from a single excitation point. Their interaction with defects produces measurable changes in amplitude, phase, propagation velocity, attenuation, reflection, transmission, scattering, and mode conversion, enabling continuous monitoring and quantitative defect assessment with relatively few sensors.[2,6–8]

Despite these advantages, guided-wave inspection remains challenging because the complex rail geometry supports multiple dispersive longitudinal, flexural, torsional, and shear-horizontal modes. Multimodal propagation, mode conversion, defect scattering, and varying boundary conditions introduced by welds, sleepers, and fastening systems complicate signal interpretation and quantitative damage characterization. Consequently, identifying guided-wave modes that simultaneously provide high defect sensitivity and long propagation distances remains an active area of research.[3,6,9]

Extensive numerical studies using the Semi-Analytical Finite Element (SAFE) method, three-dimensional finite-element (FE) models, and hybrid SAFE–FE formulations have significantly improved the understanding of guided-wave propagation in rails. These approaches enable prediction of dispersion characteristics, mode shapes, energy distribution, and wave–defect interactions, while also supporting the optimization of transducer placement, excitation frequency, and mode selection before experimental validation.[2,3,6,10,11]

Significant advances have also been made in guided-wave excitation techniques. Piezoelectric transducers remain the most widely used owing to their simplicity and high signal-to-noise ratio, although they require physical contact and are sensitive to coupling conditions. Electromagnetic acoustic transducers (EMATs) enable non-contact inspection of ferromagnetic rails but generally produce weaker signals. More recently, laser ultrasonics have enabled fully non-contact excitation and detection with high spatial resolution, making them particularly suitable for detecting small surface-breaking cracks and weld defects while facilitating integration with advanced signal processing and numerical modelling for quantitative rail inspection.[4,8]

The rapid advancement of sensing technologies and digital data acquisition has substantially increased the volume and complexity of ultrasonic data generated during structural health monitoring and nondestructive testing. Consequently, machine learning (ML) has emerged as a powerful tool for automatically extracting damage-sensitive information from complex ultrasonic datasets, enabling improved defect detection, localization, material characterization, and quantitative damage assessment. Unlike conventional threshold-based signal processing, ML algorithms learn nonlinear relationships directly from experimental data, providing greater robustness and accuracy in ultrasonic damage characterization.[12–15]

Most ultrasonic nondestructive testing applications employ supervised learning because of the availability of labeled datasets. Widely adopted algorithms include Support Vector Machines (SVM), Artificial Neural Networks (ANN), Decision Trees (DT), Random Forests (RF), Gaussian Process Regression (GPR), Gradient Boosting Regression (GBR), and Extreme Gradient Boosting (XGBoost). These models have demonstrated excellent performance in crack localization, defect identification, material characterization, and crack-depth

prediction, with ensemble methods such as RF and GBR showing superior robustness against measurement noise and complex wave–defect interactions.[16-20]

A key component of ML-based ultrasonic inspection is feature engineering, in which raw ultrasonic signals are preprocessed through filtering, normalization, time alignment, and wave-packet extraction before computing damage-sensitive descriptors in the time, frequency, and time–frequency domains. Representative features include peak amplitude, time-of-flight (ToF), root mean square (RMS), signal energy, attenuation coefficient and wavelet energy. These physically meaningful features capture crack-induced changes in wave propagation while reducing data dimensionality and combining multiple feature domains has consistently been shown to improve quantitative damage prediction compared with single-feature approaches.[12,13,19]

Recent advances in deep learning have further improved ultrasonic inspection by automatically learning hierarchical representations from raw guided-wave signals, spectrograms, wavelet scalograms, and ultrasonic images. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and transformer architectures have achieved excellent performance in defect detection and localization. However, these models generally require large, labeled datasets, substantial computational resources, and careful hyperparameter optimization, while their black-box nature limits engineering interpretability in safety-critical railway applications.[13,16]

To overcome these limitations, recent research has increasingly focused on explainable artificial intelligence (XAI) and hybrid ML frameworks that integrate engineering knowledge with data-driven models. Techniques such as SHAP, permutation importance, and sensitivity analysis improve model transparency by identifying the ultrasonic features that contribute most significantly to damage prediction, while also enhancing generalization and reducing dependence on extensive experimental datasets, particularly in railway applications where comprehensive labeled data are difficult and expensive to obtain.[20-24]

Artificial intelligence (AI) has also become a powerful tool for railway structural health monitoring by enabling automated interpretation of complex guided-wave signals. Recent studies have successfully applied supervised and deep learning algorithms, including CNNs, autoencoders, and probabilistic deep learning models, to detect transverse defects, weld defects, rolling contact fatigue, and rail-base cracks from guided-wave signals and their time–frequency representations.[25–29] Beyond defect classification, AI has increasingly been applied to quantitative damage characterization and uncertainty-aware assessment. Zhang *et al.*[28] integrated ultrasonic guided waves with Bayesian deep learning for simultaneous rail damage identification and uncertainty quantification, while Lu *et al.*[29] combined FE acoustic-field modeling with deep learning to reconstruct three-dimensional rail defect geometry from ultrasonic B-scan data, demonstrating improved reconstruction accuracy with reduced experimental data requirements.

Recent studies have further extended AI to continuous crack-depth estimation. Lv *et al.*[30] combined guided waves, wavelet time–frequency analysis, and deep learning models, including MobileNet and GoogLeNet, to estimate crack depth in turnout switch rails with recognition accuracies approaching 95%. Nevertheless, most existing approaches formulate crack assessment as an image-classification problem rather than continuous regression, and their reliance on large, labelled datasets and limited physical interpretability restricts practical deployment in railway inspection.[30]

Among physics-informed approaches, Physics-Informed Neural Networks (PINNs) have emerged as a promising framework by incorporating governing partial differential equations and boundary conditions directly into the learning process. Shukla *et al.* demonstrated the capability of PINNs for ultrasonic nondestructive testing by identifying and quantifying surface-breaking cracks from laser-generated ultrasonic surface waves, showing that embedding the acoustic wave equation enables accurate estimation of local wave-velocity variations associated with crack formation.[31–33]

Subsequent studies have extended physics-informed learning to ultrasonic wave propagation, material characterization, and guided-wave structural health monitoring. By integrating FE simulations with limited experimental data, PINNs have been used to reconstruct ultrasonic wavefields, estimate material properties, and characterize defects in metallic and composite structures. More recently, defect-parameterized PINNs have enabled accurate characterization of sub-millimeter defects from laser ultrasonic wavefields while significantly reducing training data requirements, demonstrating improved prediction accuracy, robustness, and generalization over purely data-driven approaches.[32–34]

In parallel, physics-guided machine learning has emerged as a complementary approach that employs physically meaningful ultrasonic features, such as ToF, attenuation, signal energy, spectral centroid, wavelet energy, reflection coefficients, and nonlinear ultrasonic parameters, instead of explicitly embedding governing equations into the learning process. Combined with

FE-generated and experimental datasets, these features enable conventional regression models, including GBR, RF, and GPR, to achieve accurate and interpretable quantitative damage assessment under limited data conditions.[33–35]

From a broader acoustic perspective, the observed reduction in Rayleigh-wave amplitude with increasing crack depth can be interpreted as a progressive redistribution of incident wave energy through reflection, scattering, mode conversion, and other wave–defect interactions. Related studies on architected acoustic systems have demonstrated that structural configuration, resonator coupling, and exceptional-point phenomena can strongly modify sound absorption and energy transport[36,37], while recent developments in non-integer-dimensional architected materials and acousto-mechanical metamaterials further demonstrate how structural architecture and multiphysical coupling can govern acoustic-energy localization, transmission, and dissipation.[38,39] Although these systems differ from cracked rail structures, they provide a complementary physical perspective supporting the present observation that increasing structural discontinuity can redistribute propagating wave energy and thereby alter amplitude- and energy-based ultrasonic descriptors.

Recent advances further demonstrate the expanding role of AI and machine learning across engineering analysis and condition-assessment problems. AI-enhanced modeling has been investigated for complex power-system behavior[40], while deep-learning approaches have been applied to the optimal design of composite structures[41] and deep-feature-fusion methods to automated detection of mechanical components.[42] Collectively, these studies demonstrate the broader capability of data-driven methods to extract complex relationships from engineering data and support prediction, design, and automated assessment. However, their application domains differ substantially from ultrasonic rail inspection, highlighting the need for domain-specific, physically interpretable ML frameworks that integrate ultrasonic wave characteristics for quantitative crack assessment.

Despite these advances, the application of physics-informed and physics-guided learning to continuous crack-depth estimation in rail-track structures remains limited. Most existing studies focus on simplified laboratory specimens, material-property estimation, wavefield reconstruction, or defect classification rather than quantitative crack-depth regression under realistic railway conditions. Moreover, few studies have combined guided-wave measurements with interpretable machine-learning models for direct crack-depth prediction. To address this gap, the present study proposes a physics-guided machine-learning framework for continuous ultrasonic crack-depth quantification in rail-track structures. The proposed approach employs physically meaningful ultrasonic features to train interpretable regression models for accurate and baseline-free crack-depth prediction. The term physics-guided is used here specifically in the feature-level sense: physical knowledge of Rayleigh-wave propagation and wave–crack interaction informs the selection of damage-sensitive descriptors, whereas no governing wave equation is imposed as a constraint on the regression loss function. The present framework is therefore distinguished from equation-level physics-informed ML and PINNs.

The novelty of this work is as follows.

(i) The study evaluates continuous-valued crack-depth regression using laser-ultrasonic measurements obtained from physical rail specimens containing controlled surface-breaking defects up to 20 mm deep.
(ii) A feature-level physics-guided framework is developed using damage-sensitive descriptors associated with Rayleigh-wave amplitude, energy, propagation, attenuation, and time–frequency behaviour.
(iii) Specimen-wise data partitioning and grouped cross-validation are employed so that repeated measurements from the same physical specimen do not occur simultaneously in training and validation/test subsets.
(iv) Multiple nonlinear regression models are systematically compared through prediction performance, feature importance, permutation importance, ablation, correlation analysis, and uncertainty quantification.
(v) Simple linear-regression baselines are additionally evaluated to determine whether model complexity provides predictive benefit over dominant individual ultrasonic descriptors.

In this context, the present study proposes a physics-guided ML framework for continuous crack-depth quantification in rail-track structures using ultrasonic guided waves. The proposed approach uses experimental measurements and employs physically meaningful ultrasonic features to train interpretable regression models for baseline-free crack-depth prediction. Unlike existing studies that primarily focus on defect detection or classification, this work addresses continuous crack-depth estimation while using experimental data to improve prediction accuracy, generalization, and engineering interpretability. The proposed framework provides a robust and practical solution for intelligent ultrasonic structural health monitoring of railway infrastructure.

2. Material properties and Rayleigh wave velocity

The rail-track specimen investigated in this study was manufactured from high-strength rail steel with a density of 7799 kg/m^3, Young's modulus of 212 GPa, and Poisson's ratio of 0.2866. From these elastic constants, the corresponding Lamé constants were calculated as λ = 110.7 GPa and μ = 82.4 GPa. Using these material properties, the bulk longitudinal and shear wave velocities were first determined, followed by the calculation of the Rayleigh surface wave velocity employed throughout the experimental investigations. The bulk wave velocities are obtained from the Lamé constants as:

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)} \tag{1}$$

$$\mu = \frac{E}{2(1+\nu)} \tag{2}$$

$$C_L = \sqrt{\frac{\lambda+2\mu}{\rho}} \tag{3}$$

$$C_T = \sqrt{\frac{\mu}{\rho}} \tag{4}$$

The Rayleigh wave velocity is then estimated from the shear-wave velocity using the well-established approximation:

$$C_R = \left(\frac{0.87+1.12\nu}{1+\nu}\right) C_T \tag{5}$$

For the present rail-track specimen, these equations yield the longitudinal wave velocity of 5943.49 m/s, shear wave velocity of 3250.46 m/s, and Rayleigh wave velocity of 3008.91 m/s. The calculated Rayleigh wave velocity is essentially independent of excitation frequency, confirming the non-dispersive behavior of Rayleigh waves in a homogeneous isotropic elastic medium. This property enables accurate selection of the excitation angle, wavelength, and inspection frequency for laser ultrasonic experiments. The actual image of the rail track specimen with crack and line-arrayed laser pattern is shown in Figure 1A and 1B. The principal experimental parameters relevant to the present study are described in section 4, while additional details of the laser-ultrasonic system and its development are available in the author's previous studies.[43–47]

3. Overview of the proposed framework

The proposed framework integrates experimental ultrasonic guided-wave measurements, physics-guided feature extraction, and ML regression to enable accurate, robust, and interpretable crack-depth quantification in rail-track structures. Unlike conventional ultrasonic inspection methods that primarily focus on defect detection or classification, the proposed approach formulates crack-depth estimation as a continuous regression problem by learning the relationship between physically meaningful ultrasonic features and crack geometry. The overall workflow is illustrated in Figure 2.

The framework consists of four sequential stages: (i) experimental ultrasonic testing of rail specimens with different crack depths, (ii) signal preprocessing and physics-guided feature extraction, (iii) development of a physics-guided machine learning regression model, and (iv) validation using unseen experimental data. Experimental measurements provide labeled ultrasonic datasets representing different crack conditions and form the basis for developing and evaluating the proposed learning framework.

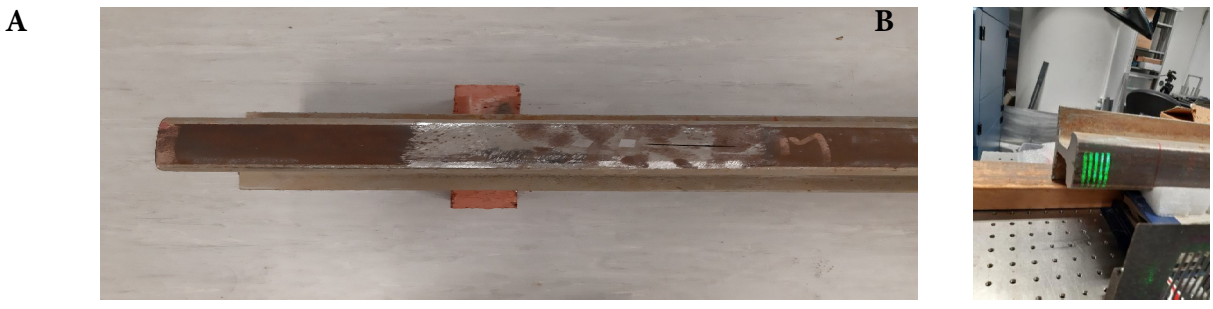


Figure 1. Rail-track specimens: (A) specimen with surface crack; and (B) specimen with incident line arrayed pattern of laser beam for generating narrowband Rayleigh-guided wave.

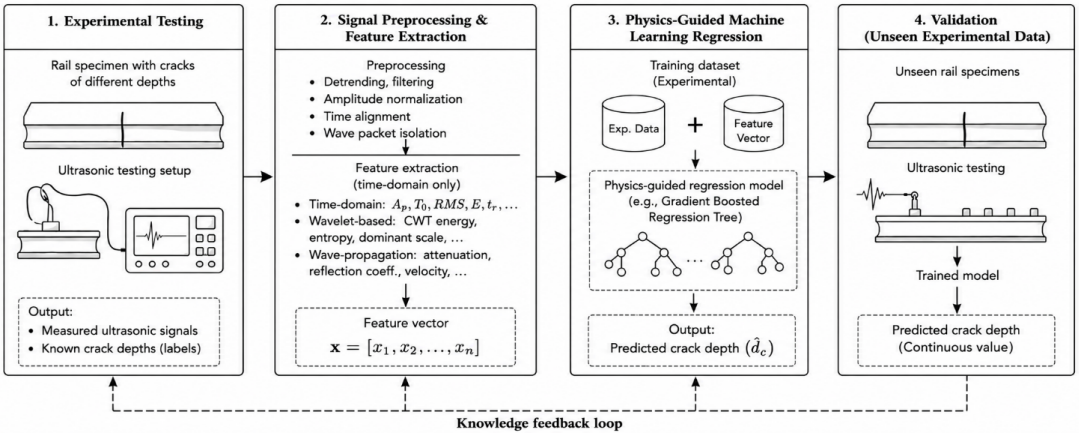


Figure 2. Overall workflow of the proposed physics guided wave machine learning framework for the crack depth quantification in rail track structures

The acquired ultrasonic signals are preprocessed through detrending, filtering, normalization, time alignment, and wave-packet isolation before extracting damage-sensitive features in the time, frequency, time–frequency, and wave-propagation domains. Representative features include peak amplitude, ToF, RMS, signal energy, attenuation coefficient, wavelet energy, and wavelet entropy. These physically meaningful descriptors characterize crack-induced changes in ultrasonic wave propagation and form the input feature vector for model development. The extracted feature vectors, together with the corresponding crack-depth labels, are used to train a Gradient Boosted Regression Tree (GBRT) model for continuous crack-depth prediction. By incorporating physically interpretable ultrasonic descriptors rather than relying directly on raw signals, the framework enables interpretable crack-depth regression and facilitates systematic evaluation of the relative contribution and redundancy of different wave-propagation features.

Finally, the trained model is validated using independent experimental specimens that are excluded from the training process. Model performance is evaluated using the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), together with feature-importance analysis to identify the ultrasonic parameters that contribute most significantly to crack-depth estimation. Overall, the proposed framework provides an accurate, interpretable, and scalable methodology for quantitative crack-depth assessment and intelligent structural health monitoring of railway infrastructure.

4. Experimental ultrasonic measurements and signal preprocessing

Experimental validation was performed using rail-track specimens containing artificial surface cracks with depths ranging from 0 to 20 mm. The artificial surface-breaking defects were introduced by a thin cutting disc, producing approximately 30 mm wide, parallel discontinuities (to wave propagation direction) with nominal depths of 2–20 mm. Ultrasonic guided-wave measurements were acquired using fixed excitation frequency, sensor spacing, coupling conditions, and acquisition parameters to ensure experimental consistency. Each crack depth was measured repeatedly to improve statistical reliability. Prior to feature extraction, the acquired ultrasonic signals were pre-processed to reduce measurement variability while preserving crack-induced wave characteristics. As illustrated in **Figure 3**, prior to feature extraction, the acquired ultrasonic signals were preprocessed by removing the DC offset, resampling signals with nonuniform time increments when required, applying band-pass filtering, and isolating the Rayleigh-wave analysis window. The amplitude scale of the filtered waveform was retained during extraction of amplitude-sensitive descriptors, including peak amplitude, peak-to-peak amplitude, RMS, signal energy, and attenuation, so that crack-induced amplitude variations were preserved. Amplitude normalization was applied only to a separate copy of the direct-wave packet used for wavelet-based feature calculation, thereby reducing sensitivity of the wavelet descriptors to absolute signal magnitude while preserving amplitude information in the corresponding physical features.

It should be emphasized that amplitude normalization was not applied to the waveform used for extracting amplitude-sensitive features. Peak amplitude, peak-to-peak amplitude, RMS, signal energy, and attenuation were evaluated directly from the filtered direct Rayleigh-wave packet, thereby preserving the amplitude reduction associated with increasing crack depth. For the wavelet-

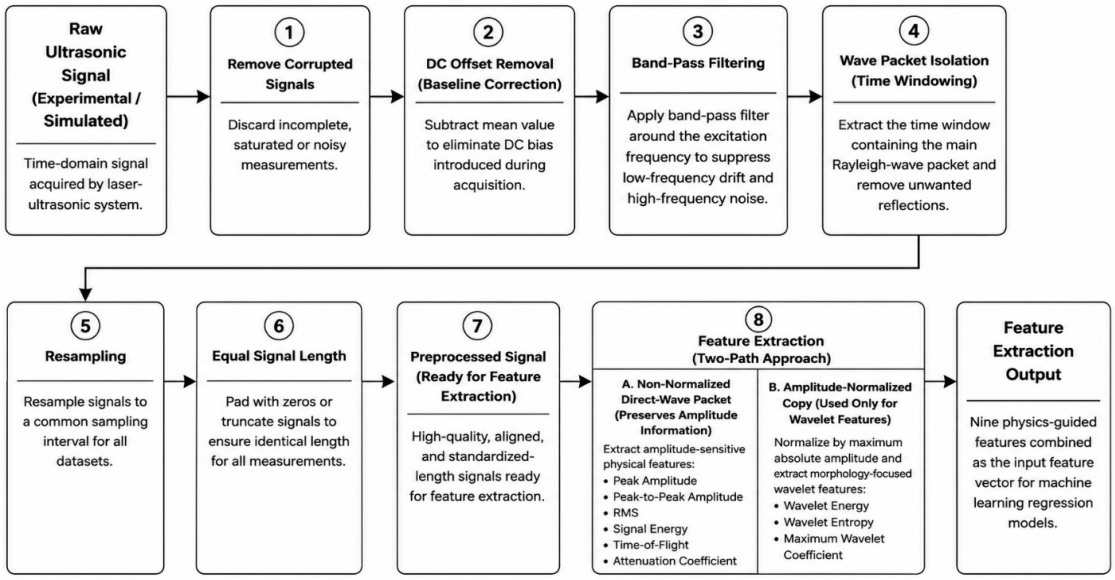


Figure 3. Block diagram of the signal pre-processing steps adopted in the proposed framework

based descriptors only, a separately normalized waveform was generated by dividing the direct-wave packet by its maximum absolute amplitude. This normalization was introduced to characterize waveform morphology and time–frequency redistribution independently of absolute signal magnitude.

The experimental measurements were performed on rail-track specimens using a fully non-contact laser-ultrasonic pitch–catch configuration previously developed by the authors.[43-47] Narrowband Rayleigh surface waves were generated using an optical line-array excitation arrangement and detected non-contactly using a scanning laser Doppler vibrometer. The line-array configuration was employed to preferentially generate narrowband Rayleigh waves while limiting the contribution of unwanted elastic-wave modes. Prior to testing, the naturally rusted rail surfaces were mechanically cleaned using a hand grinder followed by fine sandpaper to improve the quality and signal-to-noise ratio of the optical measurements.

The ultrasonic excitation was centred at 500 kHz with approximately 6.5 cycles, corresponding to a nominal Rayleigh wavelength of approximately 6.02 mm based on the Rayleigh-wave velocity reported in section 2. A fixed source–receiver separation of 60 mm was maintained during the measurements, corresponding to an expected direct Rayleigh-wave arrival time of approximately 19.94 μs. The 60-mm separation was adopted based on the previously established experimental configuration to provide adequate signal quality while reducing interference between the direct Rayleigh-wave packet and unwanted boundary reflections.[43-47]

The experimental database comprised 13 rail specimens, including three healthy specimens and ten specimens containing controlled surface-breaking defects with nominal depths ranging from 2 to 20 mm in 2-mm increments. Twenty-five repeated ultrasonic measurements were obtained from each specimen, resulting in 325 signal observations from 13 physically independent specimens. The repeated measurements associated with each specimen were retained as a group during subsequent machine-learning partitioning to prevent specimen-level data leakage. The machined/artificial surface-breaking defects used in the present experiments reproduce controlled geometric discontinuities but do not reproduce the roughness, closure, branching, residual stress and irregular morphology of fatigue-generated service cracks.

A fourth-order Butterworth band-pass filter centred on the excitation frequency (400–600 kHz for a 500 kHz excitation) was implemented using zero-phase forward–backward filtering to suppress low-frequency drift and high-frequency noise without introducing phase distortion. Time alignment was performed using the arrival of the direct wave packet as a common reference, while wave-packet isolation removed unwanted boundary reflections. Finally, all signals were resampled and standardized to identical lengths, ensuring a consistent dataset for subsequent feature extraction and machine learning analysis.

Following preprocessing, physics-guided damage-sensitive features were extracted from the ultrasonic waveforms instead of using raw signals directly. This feature-engineering approach reduces data dimensionality while preserving the physical information associated with wave attenuation, scattering, propagation delay, and energy redistribution caused by crack growth. The final feature vector comprised nine parameters: peak amplitude, peak-to-peak amplitude, RMS, signal energy, ToF, attenuation coefficient, wavelet energy, wavelet entropy, and maximum wavelet coefficient, which were subsequently used to train the regression model. The calculation of each of these features is presented below.

(a) Peak amplitude

The peak amplitude represents the maximum absolute value of the received ultrasonic signal and decreases as crack depth increases due to wave scattering and attenuation.

$$A_p = \max\left(\left|x_i\right|\right) \tag{6}$$

where x_i denotes the signal amplitude at the i^{th} sampling point.

(b) Peak-to-peak amplitude

The peak-to-peak amplitude quantifies the total dynamic range of the received waveform and is sensitive to energy loss caused by structural discontinuities.

$$A_{pp} = x_{\max} - x_{\min} \tag{7}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum signal amplitudes, respectively.

(c) Root mean square (RMS)

The RMS value provides a measure of the effective signal amplitude and reflects the average ultrasonic energy transmitted through the specimen.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{8}$$

where N is the total number of samples.

(d) Signal energy

Signal energy measures the total energy contained within the ultrasonic waveform and decreases as crack depth increases due to wave attenuation and scattering.

$$E = \sum_{i=1}^{N} x_i^2 \tag{9}$$

(e) Time-of-flight (ToF)

The ToF represents the arrival time of the received ultrasonic wave packet and is influenced by changes in propagation path caused by the presence of cracks.

$$ToF = t_r - t_e \tag{10}$$

where t_e and t_r denote the excitation and arrival times, respectively.

(f) Attenuation coefficient

The attenuation coefficient quantifies the reduction in ultrasonic wave amplitude during propagation through the cracked rail.

$$\alpha = \frac{1}{L}\ln\left(\frac{A_0}{A}\right) \tag{11}$$

where A_0 denotes the peak envelope amplitude of the direct Rayleigh-wave packet extracted from the same ultrasonic measurement, A denotes the maximum envelope amplitude of the subsequent reflected/scattered response, and L is the source–receiver distance. Because both A_0 and A are obtained from the same measurement, the attenuation descriptor does not require a separately recorded healthy-state reference signal.

(g) Wavelet energy

Continuous Wavelet Transform (CWT) was employed to analyse localized variations in the ultrasonic waveform. The wavelet energy represents the total energy contained within the wavelet coefficients and reflects localized scattering caused by crack growth.

$$WE = \sum_a \sum_b \left|W\left(a,b\right)\right|^2 \tag{12}$$

where $W\left(a,b\right)$ denotes the wavelet coefficient at scale a and translation b.

(h) Wavelet entropy

Wavelet entropy measures the complexity of the wavelet energy distribution and increases as the ultrasonic waveform becomes more irregular owing to crack-induced scattering.

$$WE_n = -\sum_{i=1}^{M} P_i \ln\left(P_i\right) \tag{13}$$

where

$$P_i = \frac{WE_i}{\sum_{i=1}^{M} WE_i} \tag{14}$$

is the normalized wavelet energy of the i^{th} decomposition level.

(i) Maximum wavelet coefficient

The maximum wavelet coefficient represents the highest localized wavelet response and provides an indication of the strongest wave–crack interaction.

$$W_{\max} = \max\left(\left|W\left(a,b\right)\right|\right) \tag{15}$$

5. Feature vector for crack depth detection

The extracted feature vector is expressed as

$$\mathbf{X} = \left[A_p, A_{pp}, RMS, E, ToF, \alpha, WE, WE_n, W_{\max}\right] \tag{16}$$

where A_p is the peak amplitude, A_{pp} is the peak-to-peak amplitude, RMS is the root mean square, E is the signal energy, ToF is the time-of-flight, α is the attenuation coefficient, WE is the wavelet energy, WE_n is the wavelet entropy, and $W_{\max}$ is the maximum wavelet coefficient. The resulting feature vector serves as the input to the ML regression model.

The target output is defined as:

$$y = d_c \tag{17}$$

where d_c denotes the crack depth.

For model development, the dataset was partitioned on a specimen-wise basis to prevent information leakage between the training and testing sets. The regression models were trained using measurements from the training specimens and evaluated on completely unseen specimens. Model performance was assessed using the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) given as follows

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{18}$$

Root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum(y_i - \hat{y}_i)^2} \tag{19}$$

Mean absolute error (MAE):

$$\mathrm{MAE} = \frac{1}{N}\sum|\, y_i - \hat{y}_i| \tag{20}$$

These metrics quantify prediction accuracy for continuous crack-depth estimation.

6. Database structure

A total of 13 rail-track specimens is considered, and they were supplied by Hong Kong MTR corporation LTD, Hong Kong. Three specimens are maintained in the healthy condition with no cracks, while the remaining ten specimens contain artificially induced cracks with depths increasing from 2 mm to 20 mm in increments of 2 mm. Each specimen is tested 25 times under identical measurement conditions to assess repeatability and generate a sufficiently large pool of ultrasonic data. Each individual ultrasonic measurement forms one row in the database. Although the database contains 325 individual ultrasonic measurements, the number of physically independent experimental specimens is 13. Consequently, the repeated measurements primarily characterize measurement repeatability rather than increasing the number of independent structural specimens.

The nine physics-guided damage features form the predictor variables, while crack depth is used as the regression target. Thus, the final ML dataset has:

$$\mathbf{X} \in \mathbb{R}^{325\times 9}$$

where the nine columns correspond to the extracted damage features, and

$$\mathbf{y} \in \mathbb{R}^{325\times 1}$$

contains the associated crack-depth labels. Here, specimen and test identifiers are retained as metadata and not as mere predictor variables.

7. Results and discussion

Figure 4 compares the normalized Rayleigh-wave signals acquired from rail specimens containing different crack depths. Note that, amplitude normalization was applied only for visualization of waveform morphology in Figure 4. Quantitative amplitude-dependent features were extracted from the filtered but non-normalized waveforms. Amplitude-dependent ultrasonic descriptors were calculated from the unnormalized filtered waveform. Feature scaling was subsequently performed on the extracted predictor variables for ML algorithms requiring standardized inputs.

The direct wave packet, arriving at approximately 20 μs, exhibits a gradual reduction in amplitude with increasing crack depth due to increased scattering and attenuation caused by the crack. In contrast, the later reflected/scattered wave packet (approximately 40–50 μs) becomes progressively more pronounced as crack depth increases, indicating enhanced wave reflection from the defect. The arrival time of the direct wave remains nearly unchanged because the propagation distance and material properties for all specimens are roughly constant, while the observed amplitude variations demonstrate the increasing

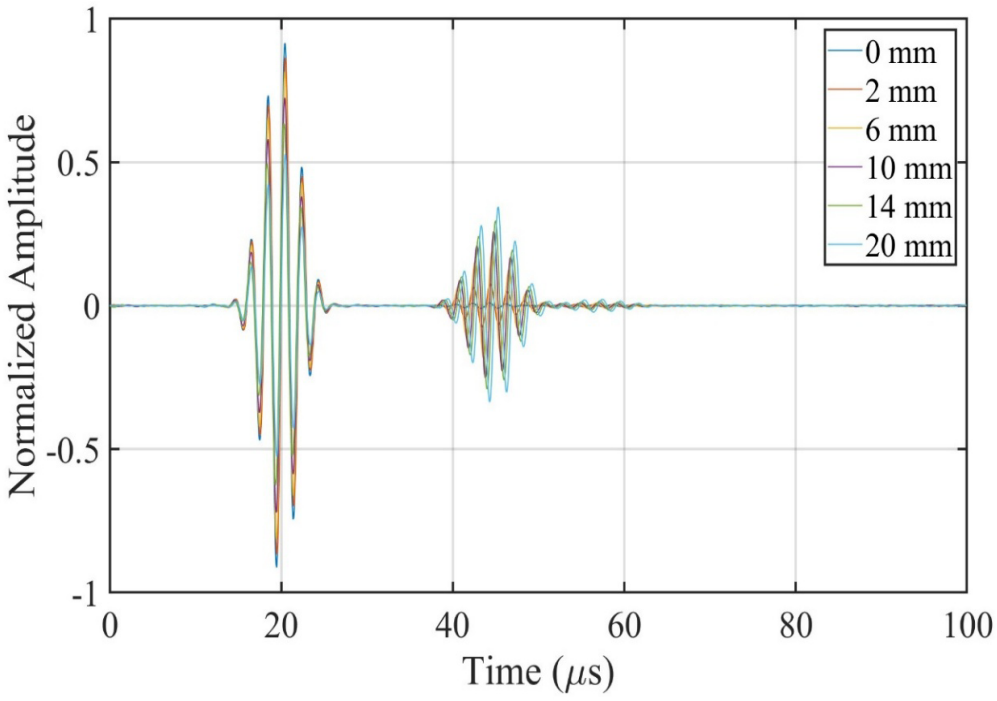


Figure 4. Time domain signals captured for different crack depths

interaction between the ultrasonic wave and the crack. These results confirm that crack depth primarily influences the wave amplitude and energy rather than the propagation velocity, providing a reliable basis for quantitative damage characterization using physics-guided ultrasonic features.

Figure 5 illustrates the variation of the extracted physics-guided ultrasonic features with increasing crack depth from 0 to 20 mm. The amplitude-based features, including peak amplitude, peak-to-peak amplitude, RMS, and signal energy, exhibit a clear monotonic decrease as crack depth increases. Specifically, the peak amplitude decreases from approximately 0.92 to 0.54, the peak-to-peak amplitude decreases from 1.82 to 1.06, the RMS value decreases from 0.20 to 0.12, and the signal energy reduces by nearly 67% over the investigated crack-depth range. This progressive reduction is attributed to increased scattering, reflection, and energy dissipation of the propagating Rayleigh wave as larger cracks remove more energy from the direct wave packet. The small error bars further demonstrate good measurement repeatability due to non-contact actuation and indicate that these amplitude-based parameters are highly sensitive to crack growth.

The ToF exhibits only minor fluctuations around 15.95 μs with overlapping error bars across all crack depths, indicating that crack propagation has a negligible influence on the propagation velocity of the direct Rayleigh wave. This behaviour is expected because the source–receiver distance and the bulk material properties remain unchanged, while the cracks primarily influence wave amplitude rather than propagation speed. In contrast, the attenuation coefficient increases almost monotonically from approximately 1.5 to 10.5, reflecting the progressive loss of ultrasonic energy due to enhanced scattering and reflection from deeper cracks. The strong inverse relationship between attenuation and amplitude-based features confirms that attenuation is an effective physics-guided indicator for quantitative crack-depth characterization.

The wavelet-based features exhibit comparatively different trends. Wavelet energy remains nearly constant throughout the crack-depth range, showing only minor variations within the measurement uncertainty. Similarly, the maximum wavelet coefficient exhibits negligible changes, indicating that these features contain limited additional information regarding crack progression for the present Rayleigh-wave dataset. In contrast, wavelet entropy shows a gradual increase from approximately 2.150 to 2.167 as crack depth increases, suggesting that deeper cracks produce greater waveform complexity due to increased scattering, multiple reflections, and localized mode conversions. Although the magnitude of change is relatively small, the consistent upward trend demonstrates that wavelet entropy is more sensitive to damage evolution than the other wavelet-based features.

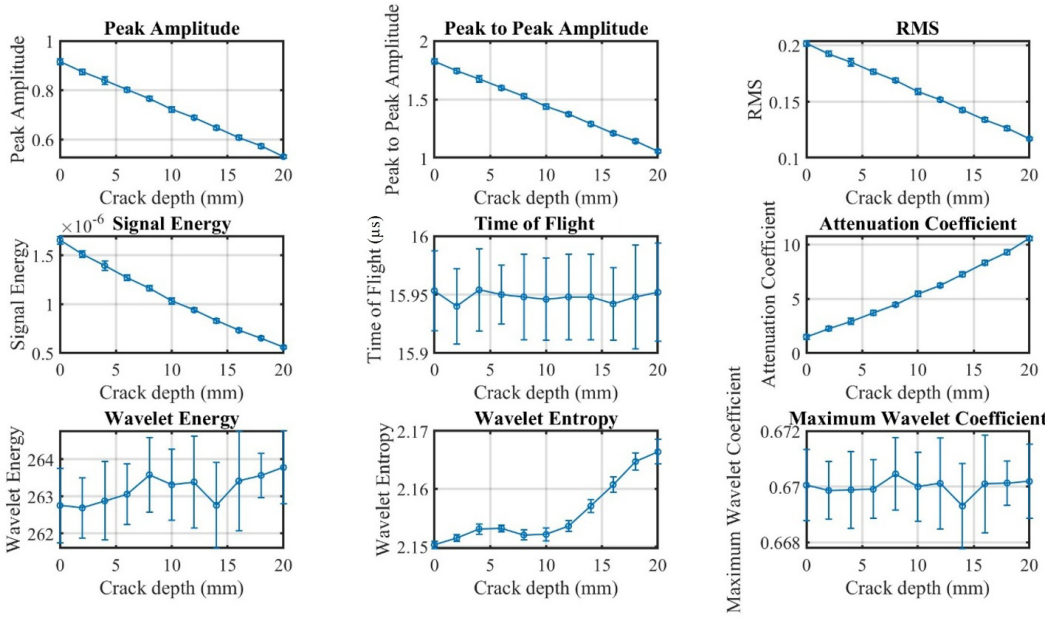


Figure 5. Variation of physics-based ultrasonic features with increasing crack depth

Overall, the extracted feature trends reveal that peak amplitude, peak-to-peak amplitude, RMS, signal energy, and attenuation coefficient exhibit the highest sensitivity to crack growth and provide clear monotonic relationships with crack depth, making them the most informative predictors for ML-based regression. Conversely, ToF, wavelet energy, and maximum wavelet coefficient contribute comparatively little to damage quantification because of their weak dependence on crack size. These observations are consistent with the subsequent feature-importance and machine learning analyses, where amplitude and attenuation related features emerge as the dominant contributors to accurate crack-depth prediction, thereby validating the proposed physics-guided feature extraction framework.

Figure 6 presents the Pearson correlation matrix illustrating the linear relationships among the extracted physics-guided ultrasonic features and the target variable (crack depth). The amplitude-based features, namely peak amplitude, peak-to-peak amplitude, RMS, and signal energy, exhibit nearly perfect positive correlations with one another (r = 0.99–1.00), indicating that they describe similar physical characteristics associated with the direct Rayleigh-wave response. These features show an almost perfect negative correlation with Crack Depth ($r \approx -0.99$ to -1.00), confirming that increasing crack depth results in progressive attenuation and energy loss of the propagating ultrasonic wave. Conversely, the attenuation coefficient exhibits an almost perfect positive correlation with crack depth (r = 0.99) and a strong negative correlation with the amplitude-based features ($r \approx$ -1.00), demonstrating that attenuation increases consistently as crack size increases.

Among the remaining features, wavelet entropy shows a strong positive correlation with crack depth (r = 0.88) and attenuation (r = 0.91), suggesting that waveform complexity increases as wave scattering becomes more pronounced in deeper cracks. In contrast, ToF exhibits negligible correlation with all other features ($|r| < 0.05$), indicating that crack growth has little influence on the propagation velocity of the direct Rayleigh wave. Similarly, wavelet energy and maximum wavelet coefficient display relatively weak correlations with crack depth (r = 0.28 and r = 0.01, respectively), implying limited sensitivity to damage progression. Overall, the correlation analysis identifies the amplitude and attenuation-based parameters as the most informative features for crack-depth estimation, which is consistent with the subsequent ML feature-importance analysis and validates their selection within the proposed physics-guided regression framework. The variables with $|r| \geq 0.95$ were treated as a redundant cluster. Multicollinearity has limited effect on the predictive capability of tree ensembles but can make individual feature-importance values unstable because importance may be shared among strongly correlated predictors.

Figure 7 compares the crack-depth prediction performance of the four ML regression models using

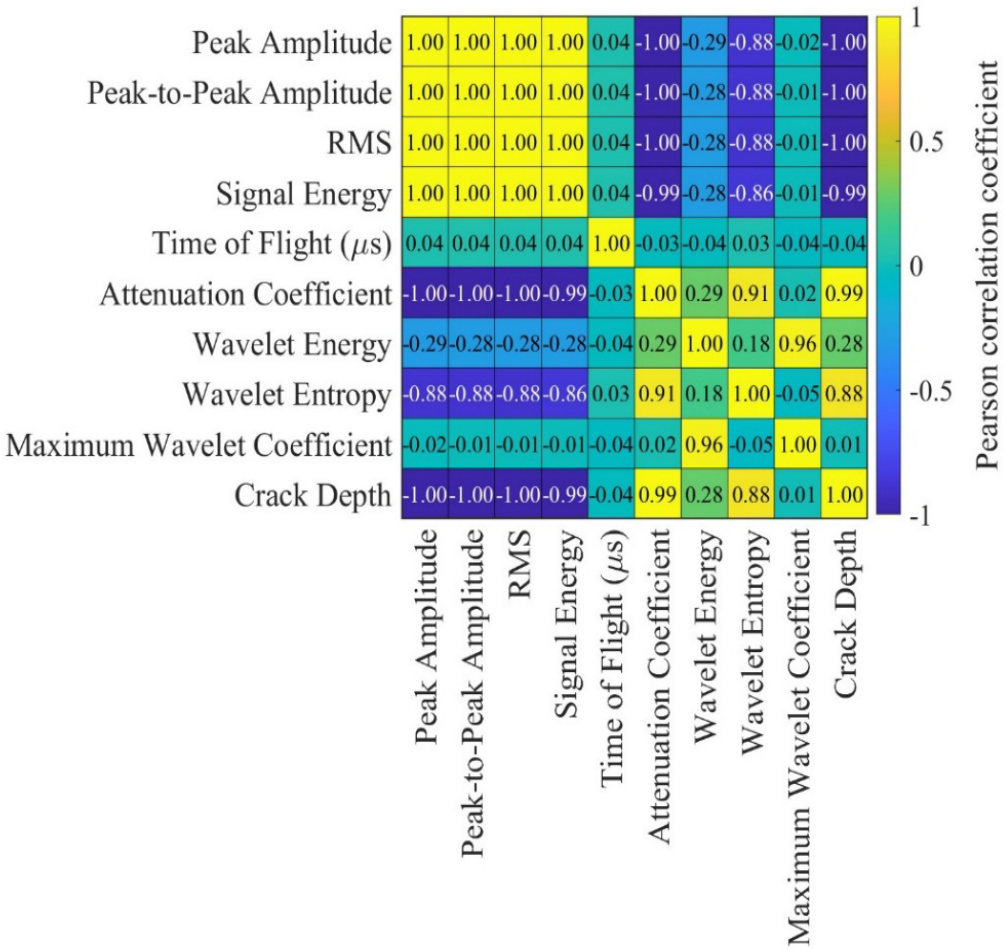


Figure 6. Pearson correlation matrix

unseen experimental specimens. The problem is formulated as continuous crack-depth regression rather than discrete damage classification. However, because the experimental database contains cracks at prescribed discrete depth increments and only one cracked specimen is available at each non-zero depth, the present study does not constitute exhaustive validation over arbitrary intermediate crack depths. Accordingly, the term continuous refers to the continuous numerical output of the regression model rather than experimentally demonstrated prediction at every possible crack depth. The dashed diagonal line represents the ideal one-to-one correspondence between the actual and predicted crack depths. The GBRT (LSBoost) and RF models exhibit excellent predictive performance, with most data points closely clustered around the ideal line. The RF model achieves the highest coefficient of determination (R^2 = 0.961) and the lowest prediction error (RMSE = 1.610 mm), followed closely by GBRT (R^2 = 0.960, RMSE = 1.623 mm). These results indicate that the ensemble tree-based models effectively capture the nonlinear relationship between the extracted physics-guided ultrasonic features and crack depth, resulting in accurate and robust quantitative damage estimation.

In comparison, the GPR model demonstrates satisfactory predictive capability (R^2 = 0.925, RMSE = 2.235 mm) but exhibits greater dispersion around the ideal prediction line, particularly for larger crack depths. The SVR model shows the weakest performance among the evaluated methods, with the lowest coefficient of determination (R^2 = 0.781) and the highest prediction error (RMSE = 3.822 mm), indicating reduced capability to model the complex nonlinear relationship between ultrasonic features and crack depth. Overall, the comparison demonstrates that ensemble learning approaches, particularly RF and GBRT (LSBoost), provide superior prediction accuracy and generalization capability for physics-guided ultrasonic crack-depth quantification in rail-track structures. The architecture and hyperparameters adopted for the four

regression models are summarized in Table 1. Fixed model configurations were retained throughout the analysis to ensure consistency between the independent specimen-wise evaluation and grouped cross-validation.

Figure 8 presents the residual distributions of the four ML regression models evaluated for ultrasonic crack-depth prediction. An ideal regression model should produce residuals that are randomly distributed around the zero-error line, indicating unbiased predictions and consistent performance across the entire crack-depth range. The GBRT (LSBoost) and RF models exhibit relatively small residuals with no pronounced systematic trend, demonstrating good agreement between the predicted and actual crack depths. Their residuals remain close to zero for most observations, suggesting stable prediction accuracy and effective generalization to unseen specimens.

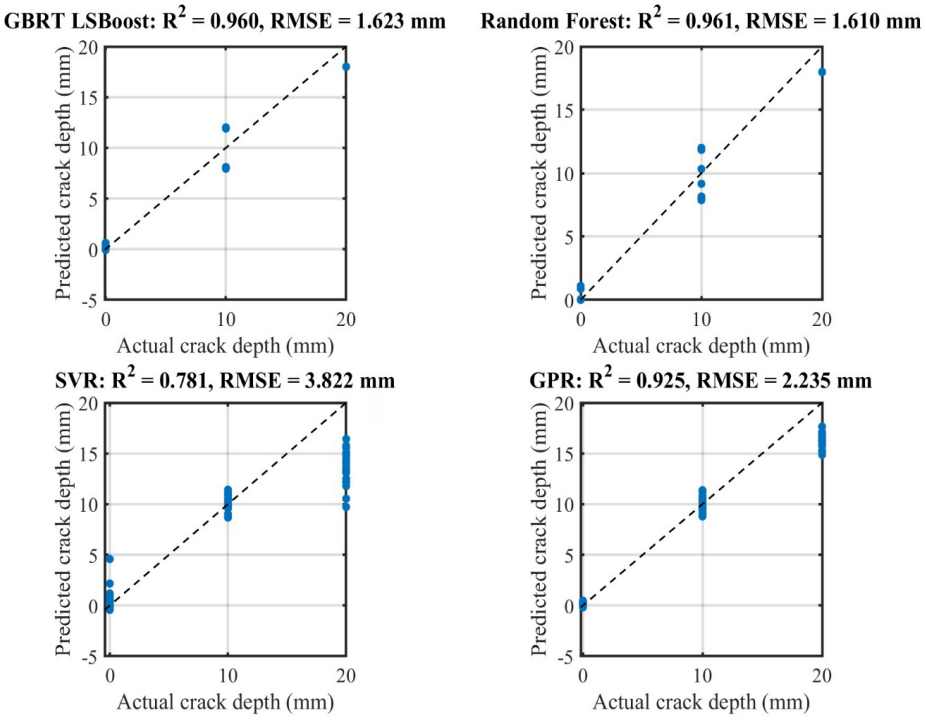


Figure 7. Crack depth prediction performance

Table 1. Regression model architecture and hyperparameters

Regression model	Architecture / Method	Hyperparameters used
GBRT-LSBoost	Gradient-boosted regression tree ensemble using least-squares boosting (LSBoost)	Number of learning cycles = 300; learning rate = 0.05; minimum leaf size = 5; maximum number of splits per tree = 20; surrogate splits = Off
Random Forest (RF)	Bagged regression-tree ensemble	Number of trees/learning cycles = 300; minimum leaf size = 3; predictors sampled per split = ceil($\sqrt{p}$) = 3 for 9 input features; surrogate splits = Off
Support Vector Regression (SVR)	ε-SVR with Gaussian (RBF) kernel	Kernel = Gaussian; kernel scale = Auto; box constraint C = 10; ε = 0.1; predictor standardization = Yes
Gaussian Process Regression (GPR)	Gaussian-process regression with ARD squared-exponential covariance kernel	Kernel = ARD squared exponential; basis/mean function = Constant; predictor standardization = Yes; fitting method = Exact; prediction method = Exact

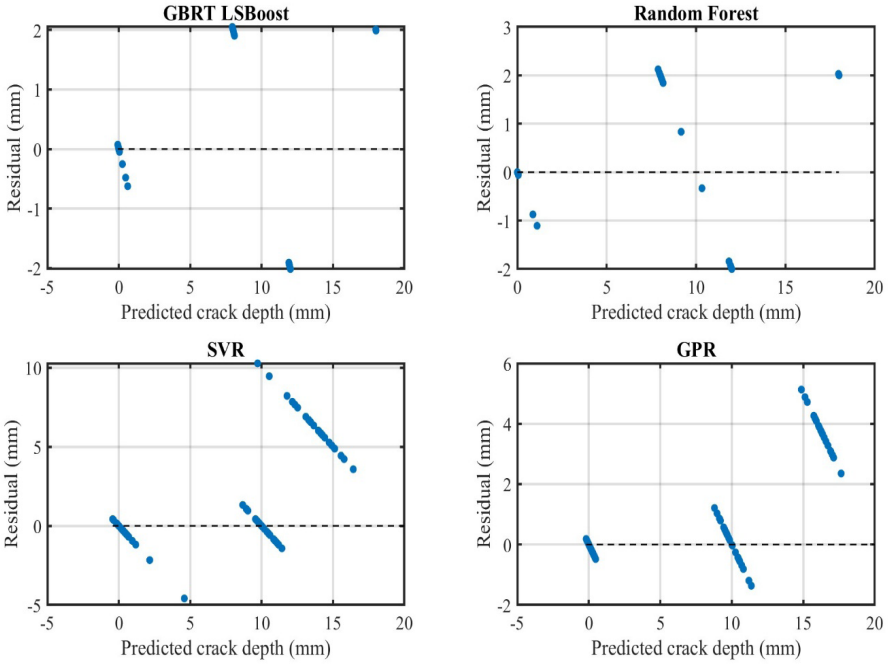


Figure 8. Residual distribution performance

In contrast, the SVR and GPR models display larger residual magnitudes and noticeable systematic patterns, particularly for specimens with larger crack depths. The increasing residuals at higher predicted crack depths indicate a tendency of these models to either underpredict or overpredict damage severity, resulting in reduced predictive accuracy. Overall, the residual analysis demonstrates that the ensemble tree-based approaches, particularly GBRT (LSBoost) and RF, provide more reliable and robust crack-depth estimation than the kernel-based regression models for the proposed physics-guided ultrasonic feature set.

Figure 9 presents the specimen-wise grouped cross-validation results for the four ML regression models, where entire specimens rather than individual measurements are excluded during validation to assess the models' ability to generalize to unseen rail samples. The GPR model achieves the lowest average cross-validation error (RMSE ≈ 1.1 mm), although it exhibits relatively large variability across different validation folds, as indicated by the wider error bars. Random forest and GBRT (LSBoost) demonstrate consistently low prediction errors (RMSE ≈ 1.5–1.7 mm) with comparatively smaller standard deviations, indicating stable and robust performance across different specimens. In contrast, SVR produces the highest average RMSE (approximately 2.0 mm) together with the largest uncertainty, suggesting lower robustness and reduced generalization capability. Overall, the grouped cross-validation confirms that the ensemble tree-based models provide reliable and consistent crack-depth prediction for unseen rail specimens, demonstrating the effectiveness of the proposed physics-guided ML framework.

To assess model robustness while preventing information leakage among repeated measurements from the same physical specimen, five-fold specimen-wise grouped cross-validation was performed. All 25 measurements from a given specimen were assigned exclusively to the same fold, such that no specimen contributed observations to both training and validation data within a fold. The 13 specimens were randomly allocated across the five folds using a fixed random seed, and each fold was used once for validation while the remaining specimens were used for model training. The same model architectures and hyperparameters were retained across all folds, and R^2, RMSE, MAE, and MAPE were calculated for each validation fold.

Figure 10 compares the normalized feature importance obtained from the GBRT and RF models, providing insight into the relative contribution of each physics-guided ultrasonic feature to crack-depth prediction. In the GBRT model, peak amplitude emerges as the most influential feature, accounting for approximately 54%

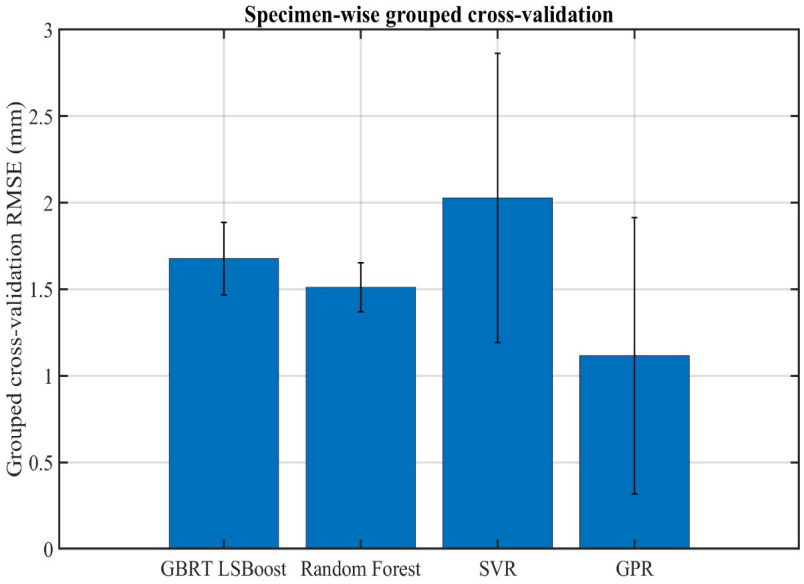


Figure 9. Specimen-wise grouped cross-validation results for the four ML regression models

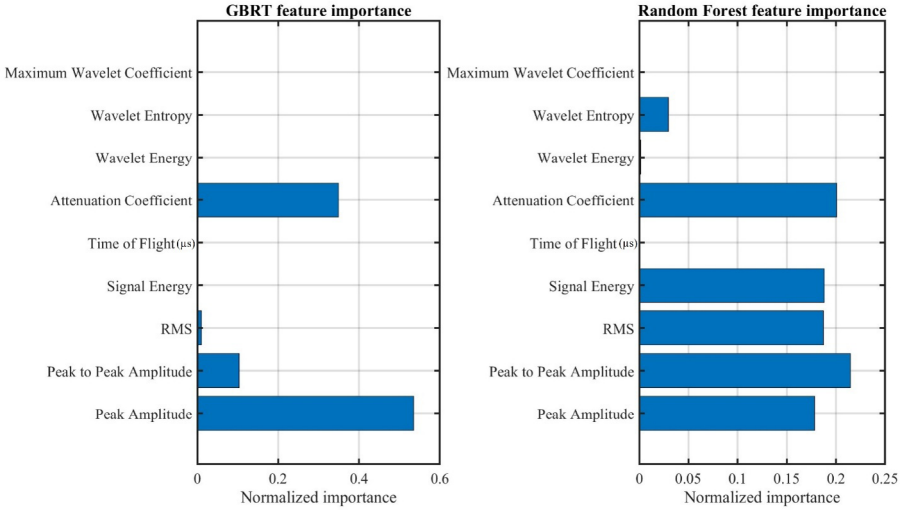


Figure 10. Feature importance from Gradient Boosted Regression Tree (GBRT)

of the total importance, followed by the attenuation coefficient (≈35%) and peak-to-peak amplitude (≈10%). The remaining features, including RMS, signal energy, ToF, wavelet energy, wavelet entropy, and maximum wavelet coefficient, contribute negligibly to the prediction process. This indicates that the boosting algorithm relies primarily on the direct-wave amplitude reduction and attenuation caused by crack-induced energy loss, which are the most sensitive indicators of damage progression.

Although GPR achieved the lowest average error during specimen-wise grouped cross-validation, its performance was comparatively lower on the independent test set. This difference is attributed to the limited number of independent physical specimens and the resulting sensitivity of GPR to the particular specimens included in each training and validation partition. In grouped cross-validation, performance is averaged across different specimen combinations, whereas the independent

evaluation uses a fixed set of completely unseen specimens (H3, C5, and C10). Consequently, the independent test set represents a specific generalization challenge that may differ from the average cross-validation condition. The more consistent independent-test performance of RF and GBRT therefore suggests greater robustness to specimen-to-specimen variability in the present dataset.

In contrast, the RF model distributes feature importance more evenly across multiple predictors. Peak-to-peak amplitude, attenuation coefficient, signal energy, RMS, and peak amplitude all exhibit comparable importance values (approximately 0.18–0.22), indicating that the ensemble of decision trees benefits from combining several complementary amplitude- and energy-related features. Wavelet entropy provides only a modest contribution, while ToF, wavelet energy, and maximum wavelet coefficient contribute minimally to the model performance. Overall, both models consistently identify amplitude- and attenuation-based features as the dominant predictors of crack depth, whereas the wavelet-based and propagation-time features provide comparatively limited additional information. These findings further validate the proposed physics-guided feature extraction strategy and are in good agreement with the feature trend and correlation analyses presented earlier.

Figure 11 presents the permutation feature importance obtained using the RF regression model. Feature importance is quantified as the increase in prediction error (RMSE) after randomly permuting the values of each feature while keeping all other features unchanged. A larger increase in RMSE indicates that the corresponding feature contributes more significantly to accurate crack-depth prediction. Among the evaluated features, peak-to-peak amplitude exhibits the highest importance, followed closely by the attenuation coefficient, signal energy, RMS, and peak amplitude. These features are directly associated with the attenuation and energy loss of the propagating Rayleigh wave caused by increasing crack depth, confirming their strong sensitivity to damage progression.

In contrast, wavelet entropy provides only a modest contribution to the prediction accuracy, while wavelet energy, ToF, and maximum wavelet coefficient exhibit negligible importance for the present dataset. The limited contribution of these features indicates that they contain comparatively less unique information beyond that already captured by the dominant amplitude- and energy-based features. Overall, the permutation analysis demonstrates that the RF model primarily relies on physically interpretable ultrasonic features related to wave attenuation and energy dissipation, supporting the effectiveness of the proposed physics-guided feature extraction framework for robust crack-depth quantification.

Figure 12 presents the feature ablation analysis performed using the RF regression model to evaluate the contribution of each physics-guided ultrasonic feature to crack-depth prediction. In this analysis, one feature is

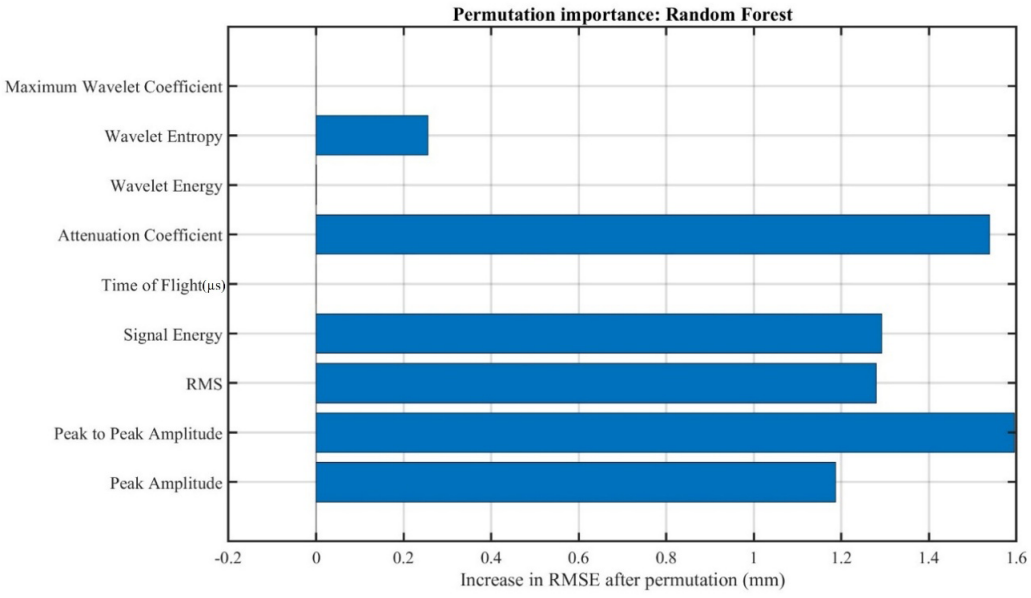


Figure 11. Permutation feature importance obtained using the Random Forest regression model

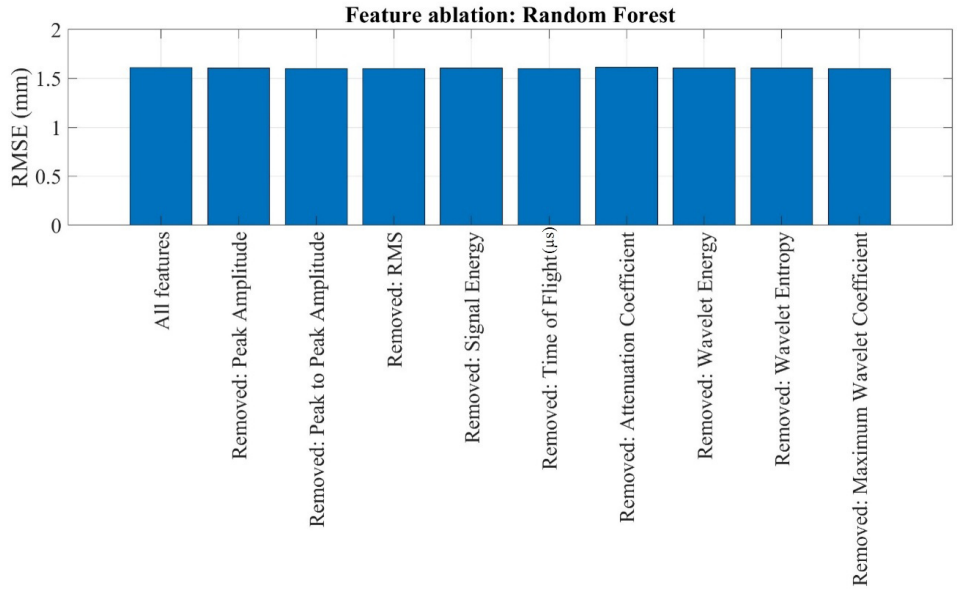


Figure 12. Feature ablation analysis using Random Forest regression

removed at a time, and the model is retrained to quantify the resulting change in prediction accuracy, expressed in terms of the RMSE. The first bar represents the baseline model using all extracted features, which achieves an RMSE of approximately 1.61 mm. Removing individual features results in only marginal changes in the prediction error, indicating that the RF model maintains robust performance even when a single feature is excluded. This behavior suggests that the selected ultrasonic features collectively provide complementary information for crack-depth estimation.

Among the evaluated features, the removal of the attenuation coefficient produces the largest increase in RMSE, highlighting its dominant contribution to the prediction process. Slight increases in RMSE are also observed when peak amplitude, peak-to-peak amplitude, signal energy, and RMS are removed, confirming that these amplitude- and energy-based features play important roles in describing crack-induced wave attenuation. In contrast, excluding ToF, wavelet energy, wavelet entropy, or maximum wavelet coefficient has a negligible effect on model performance, indicating that these features contribute relatively little additional predictive information. Overall, the ablation study confirms the robustness of the proposed feature set and further demonstrates that attenuation and amplitude-related features are the primary drivers of accurate crack-depth quantification in the proposed physics-guided machine learning framework.

Figure 13 illustrates the prediction uncertainty of the GPR model for the unseen test observations. The solid black line represents the actual crack depths, the dashed line denotes the GPR predictions, and the shaded region corresponds to the 95% prediction interval, which quantifies the confidence associated with each prediction. The predicted crack depths closely follow the actual values for the three damage levels (0, 10, and 20 mm), indicating that the GPR model successfully captures the overall nonlinear relationship between the extracted ultrasonic features and crack depth. For the healthy specimens (0 mm) and the intermediate crack depth (10 mm), the prediction intervals remain relatively narrow, reflecting higher confidence and lower predictive uncertainty.

As the crack depth increases to 20 mm, the prediction intervals become noticeably wider, indicating increased uncertainty in the model predictions. This behavior is expected because deeper cracks produce more complex wave scattering, attenuation, and nonlinear interactions, resulting in greater variability in the extracted ultrasonic features. Nevertheless, most of the predicted values remain within the 95% prediction interval, demonstrating that the GPR model provides well-calibrated uncertainty estimates while maintaining satisfactory prediction accuracy. The uncertainty analysis therefore complements the deterministic prediction results by providing confidence bounds for each estimated crack depth, which is particularly valuable for practical structural health monitoring and reliability-based decision-making.

To determine whether the predictive performance of the machine-learning models exceeded that achievable using a simple relationship between the dominant ultrasonic

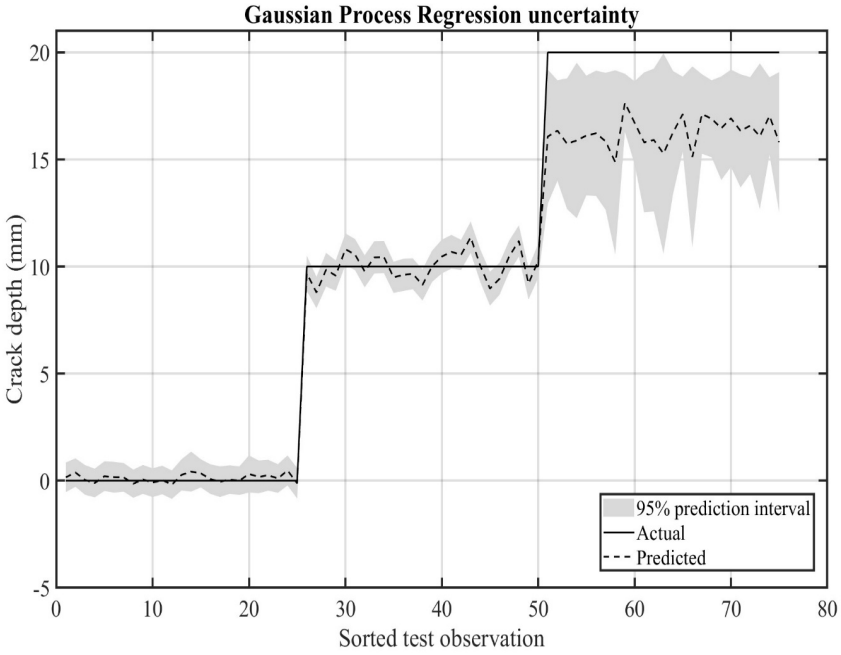


Figure 13. Prediction uncertainty of the Gaussian Process Regression

features and crack depth, ordinary least-squares regression baselines were additionally evaluated using the same specimen-wise training and independent-test partition as described in Table 2. Peak amplitude was selected as the principal single-feature baseline because it exhibited a very strong correlation with crack depth ($r = -0.997$) and was identified as the dominant predictor in the GBRT importance analysis. The peak-amplitude-only linear regression achieved $R^2 = 0.995$, RMSE = 0.563 mm, and MAE = 0.462 mm, outperforming the nine-feature ML models for the present independent test partition. This result is consistent with the highly controlled laboratory conditions, under which crack depth produced a strong and approximately linear reduction in the transmitted Rayleigh-wave amplitude. The attenuation-only model yielded $R^2 = 0.983$, RMSE = 1.071 mm, and MAE = 0.882 mm, while combining peak amplitude and attenuation provided no further improvement ($R^2 = 0.995$, RMSE = 0.565 mm, MAE = 0.467 mm), reflecting the strong redundancy between these descriptors.

Importantly, the superior performance of the single-feature baseline in the present laboratory dataset should not be interpreted as evidence that a single amplitude parameter will necessarily remain optimal under practical field conditions. In-service measurements may be influenced by environmental and operational variability, temperature, loading, rail-support conditions, surface condition, measurement positioning, material variability, and naturally propagated crack morphology, which can introduce nonlinear and interacting effects into the ultrasonic response. Under such conditions, a multi-feature nonlinear model may provide greater robustness by combining complementary information from amplitude, attenuation, propagation, and time-frequency descriptors. This hypothesis requires dedicated field validation and will be investigated in future work using a larger number of independent rails, naturally propagated cracks, and measurements acquired under varying environmental and operational conditions.

8. Limitations and future work

The present study should be interpreted in the context of several limitations. Although the experimental database contains 325 ultrasonic measurements, these measurements were obtained from only 13 independent rail specimens, with 25 repeated measurements performed on each specimen. Therefore, the repeated measurements improve the assessment of measurement repeatability but do not constitute additional independent rail specimens. Furthermore, only one cracked specimen was available at each prescribed non-zero crack depth, and the investigated defects were artificially machined under controlled laboratory conditions. Consequently, the number of independent crack-depth levels remains limited, and

Table 2. Comparison of simple regression baselines and machine-learning models for unseen-specimen crack-depth prediction

Model	Predictors	No. of predictors	R^2	RMSE (mm)	MAE (mm)	Purpose/justification
Linear regression baseline	Peak amplitude	1	0.995	0.563	0.462	Uses the dominant GBRT feature and avoids redundancy among highly correlated amplitude descriptors
Linear regression baseline	Attenuation coefficient	1	0.983	1.071	0.882	Tests whether the principal wave-attenuation descriptor alone can quantify crack depth
Multiple linear regression baseline	Peak amplitude + attenuation coefficient	2	0.995	0.565	0.467	Tests whether combining the two dominant predictors adds information beyond Peak Amplitude alone
Random Forest	All 9 features	9	0.961	1.610	1.322	Nonlinear ensemble ML model
GBRT-LSBoost	All 9 features	9	0.960	1.623	1.347	Nonlinear boosted-tree ML model
GPR	All 9 features	9	0.925	2.235	1.497	Probabilistic nonlinear regression
SVR	All 9 features	9	0.781	3.822	2.501	Kernel-based nonlinear regression

the present experiments do not capture environmental variability, rail traffic and operational loading, temperature effects, or the irregular morphology and propagation characteristics of naturally developed fatigue cracks. Although the regression models provide continuous-valued crack-depth outputs, this should not be interpreted as experimental verification of prediction accuracy at every arbitrary intermediate crack depth because only the prescribed crack-depth levels were experimentally represented. Future work should therefore expand the number of independent specimens at each crack depth, incorporate naturally propagated fatigue cracks and environmental and operational variability, and validate the proposed framework on in-service rail tracks under realistic field conditions.

9. Conclusions

This study proposed a feature-level physics-guided ML framework for ultrasonic crack-depth regression in laboratory rail specimens using laser-based experimental Rayleigh-wave measurements. Unlike conventional approaches focused on damage detection or classification, the proposed framework enables baseline-free and interpretable crack-depth regression using physically meaningful ultrasonic features.

The major findings are summarized as follows:

(i) A comprehensive dataset of healthy and cracked rail specimens (0–20 mm crack depth) was successfully developed using experimental measurements and physics-guided feature extraction.

(ii) Physics-guided ultrasonic features effectively captured crack-induced wave propagation characteristics, enabling fairly accurate quantitative crack-depth prediction especially for early-stage crack evolution.

(iii) Among the four nonlinear machine-learning models, RF achieved the best independent-test performance (R^2 = 0.961, RMSE = 1.610 mm), closely followed by GBRT-LSBoost. However, the single-feature peak-amplitude linear baseline achieved still lower prediction error (R^2 = 0.995, RMSE = 0.563 mm), indicating that the controlled laboratory dataset contains a strong approximately linear amplitude–crack-depth relationship.

(iv) Feature importance and ablation analyses confirmed the contribution of physically meaningful ultrasonic parameters, enhancing both prediction performance and engineering interpretability.

Overall, the proposed framework provides a fairly accurate, interpretable, and practical solution for intelligent ultrasonic structural health monitoring of railway infrastructure. Future work will focus on validation using in-service rail data, incorporation of environmental and operational variability, and extension to physics-informed deep learning and real-time deployment for predictive railway maintenance.

Acknowledgments

None.

Funding

None.

Conflict of interest

Faeez Masurkar is a Youth Editorial Board Member of this journal, but was not in any way involved in the editorial and peer-review process conducted for this paper, directly or indirectly. The author declares there are not any potential conflicts or competing interests with any institutes, organizations or agencies that might influence the integrity of results or objective interpretation of the submitted work.

Author contributions

This is a single-authored article.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data

Data will be made available upon request to the corresponding author.

References

1. Papaelias MP, Roberts C, Davis CL. A review on non-destructive evaluation of rails: state-of-the-art and future development. *Proc Inst Mech Eng F J Rail Rapid Transit*. 2008;222(4):367-384.

   doi: 10.1243/09544097JRRT209

2. Wilcox P, Pavlakovic B, Evans M, *et al.* Long range inspection of rail using guided waves. *AIP Conf Proc.* 2003;657:236-243.

   doi: 10.1063/1.1570142

3. Loveday PW, Long CS, Ramatlo DA. Ultrasonic guided wave monitoring of an operational rail track. *Struct Health Monit*. 2020;19(6):1666-1684.

   doi: 10.1177/1475921719893887

4. Scruby CB, Drain LE. *Laser Ultrasonics: Techniques and Applications*. New York: Routledge; 2019.

   doi: 10.1201/9780203749098

5. Rose JL. *Ultrasonic Guided Waves in Solid Media*. Cambridge: Cambridge University Press; 2014.

   doi: 10.1017/CBO9781107273610

6. Loveday PW. Guided wave inspection and monitoring of railway track. *J Nondestruct Eval*. 2012;31(4):303-309.

   doi: 10.1007/s10921-012-0145-9

7. Yang Y, Wang P, Song TL, Jiang Y, Zhou WT, Xu WL. Evaluation of the transverse crack depth of rail bottoms based on the ultrasonic guided waves of piezoelectric sensor arrays. *Sensors*. 2022;22(18):7023.

   doi: 10.3390/s22187023

8. Lian Y, Du F, Xie L, *et al.* Simulation and experimental research of V-crack testing of rail surfaces based on laser ultrasound. *Photonics*. 2024;11(10):920.

   doi: 10.3390/photonics11100920

9. Hayashi T, Song WJ, Rose JL. Guided wave dispersion curves for a bar with an arbitrary cross-section, a rod and rail example. *Ultrasonics*. 2003;41(3):175-183.

   doi: 10.1016/S0041-624X(03)00097-0

10. Setshedi II, Wilke DN, Loveday PW. Feature detection in guided wave ultrasound measurements using simulated spectrograms and generative machine learning. *NDT E Int*. 2024;143:103036.

    doi: 10.1016/j.ndteint.2024.103036

11. Xu X, Wen Z, Ni Y, Shao B, Ma X, Pan Z. Study on monitoring broken rails of heavy haul railway based on ultrasonic guided wave. *Sci Rep*. 2024;14:8667.

    doi: 10.1038/s41598-024-59328-5

12. Boris I, Barashok K, Choi Y, Choi Y, Aslam M, Lee J. Machine learning techniques in ultrasonics-based defect detection and material characterization: a comprehensive review. *Adv Mech Eng*. 2025;17(6).

    doi: 10.1177/16878132251347390

13. Yang Z, Yang H, Tian T, *et al.* A review on guided-ultrasonic-wave-based structural health monitoring: from fundamental theory to machine learning techniques. *Ultrasonics*. 2023;133:107014.

    doi: 10.1016/j.ultras.2023.107014

14. Latête T, Gauthier B, Bélanger P. Towards using convolutional neural network to locate, identify and size defects in phased array ultrasonic testing. *Ultrasonics.* 2021;115:106436.

    doi: 10.1016/j.ultras.2021.106436

15. Tang R, Zhang S, Wu W, Zhang S, Han Z. Explainable deep learning based ultrasonic guided wave pipe crack identification method. *Measurement*. 2023;206:112277.

    doi: 10.1016/j.measurement.2022.112277

16. Shen Y, Wu J, Chen J, Zhang W, Yang X, Ma H. Quantitative detection of pipeline cracks based on ultrasonic guided waves and convolutional neural network. *Sensors*. 2024;24(4):1204.

    doi: 10.3390/s24041204

17. Mehtaj N, Banerjee S. Scientific machine learning for guided wave and surface acoustic wave (SAW) propagation: PgNN, PeNN, PINN, and neural operator. *Sensors*. 2025;25(5):1401.

doi: 10.3390/s25051401

18. Casartelli A, Lomazzi L, Pinello L, Junges R, Giglio M, Cadini F. Physics-informed machine learning for ultrasonic guided wave full-field reconstruction and damage diagnosis. In: Proceedings of the 15th International Workshop on Structural Health Monitoring. Lancaster: Destech Publications, Inc.; 2025.

doi: 10.12783/shm2025/37479

19. Barshok K, Choi JI, Lee J. Deep learning-based approach for automatic defect detection in complex structures using PAUT data. *Sensors*. 2025;25(19):6128.

doi: 10.3390/s25196128

20. Xiao J, Cui F. Machine learning enhanced characterization of surface defects using ultrasonic Rayleigh waves. *NDT E Int*. 2023;140:102969.

doi: 10.1016/j.ndteint.2023.102969

21. Na Y, He Y, Deng B, *et al.* Advances of machine learning in phased array ultrasonic non-destructive testing: a review. *AI*. 2025;6(6):124.

doi: 10.3390/ai6060124

22. Vasilache MM, Tayong RB, Velisavljevic V. Comprehensive review on the use of machine learning techniques applied to the ultrasound data for the characterisation of porosity across carbon fibre reinforced polymer layers. *Appl Compos Mater*. 2025;32(4):1315-1339.

doi: 10.1007/s10443-025-10342-4

23. Bai J, Wu D, Shelley T, *et al.* A comprehensive survey on machine learning driven material defect detection. *ACM Comput Surv*. 2025;57(11):1-36.

doi: 10.1145/3730576

24. Tunukovic V, McKnight S, Pierce SG, *et al.* Application of machine learning techniques for defect detection, localisation, and sizing in ultrasonic testing of carbon fibre reinforced polymers. In: BINDT Aerospace Event 2023. Glasgow, UK; 2023. Accessed September 3, 2026. https://pureportal.strath.ac.uk/en/publications/application-of-machine-learning-techniques-for-defect-detection-l/

25. Wang Y, Miao B, Zhang Y, Huang Z, Xu S. Review on rail damage detection technologies for high-speed trains. *Appl Sci*. 2025;15(14):7725.

doi: 10.3390/app15147725

26. Wu Y, Zhu X. Rail defect detection using ultrasonic A-scan data and deep autoencoder. *Transp Res Rec*. 2023;2677(7):62-73.

doi: 10.1177/03611981221150923

27. Islam MA, Olm G. Deep learning techniques to detect rail indications from ultrasonic data for automated rail monitoring and maintenance. *Ultrasonics*. 2024;140:107314.

doi: 10.1016/j.ultras.2024.107314

28. Zhang Y, Dang DZ, Wang YW, Ni YQ. Damage identification for railway tracks using ultrasound guided wave and hybrid probabilistic deep learning. *Constr Build Mater*. 2024;418:135466.

doi: 10.1016/j.conbuildmat.2024.135466

29. Lu XL, Gao B, Woo WL, Xiao X, Zhan D, Huang C. Hybrid physics machine learning for ultrasonic field guided 3D generation and reconstruction of rail defects. *NDT E Int*. 2024;146:103174.

doi: 10.1016/j.ndteint.2024.103174

30. Lv C, Wang P, Sun H. Deep-learning-based ultrasonic guided wave detection of turnout switch rail cracks. *Sens Mater*. 2026;38(2):857-870.

doi: 10.18494/SAM5934

31. Shukla K, Di Leoni PC, Blackshire J, Sparkman D, Karniadakis GE. Physics-informed neural network for ultrasound nondestructive quantification of surface breaking cracks. *J Nondestruct Eval*. 2020;39(3):61.

doi: 10.1007/s10921-020-00705-1

32. Shukla K, Jagtap AD, Blackshire JL, Sparkman D, Karniadakis GE. A physics-informed neural network for quantifying the microstructural properties of polycrystalline nickel using ultrasound data: a promising approach for solving inverse problems. *IEEE Signal Process Mag*. 2022;39(1):68-77.

doi: 10.1109/MSP.2021.3118904

33. Pérez E, Ardıç CE, Çakıroğlu O, *et al.* Integrating AI in NDE: techniques, trends, and further directions. *NDT E Int*. 2025;156:103442.

doi: 10.1016/j.ndteint.2025.103442

34. Yang L, Liu P, Yi K, *et al.* Defect-parameterized physics-informed neural network for forward and inverse modeling of laser ultrasonic wavefield. *Ultrasonics*. 2026;161:107952.

doi: 10.1016/j.ultras.2026.107952

35. Casartelli A, Lomazzi L, Giglio M, Cadini F. Physics-informed neural network for baseline-free damage diagnosis using ultrasonic guided waves. *e-J Nondestruct Test*. 2026;31(8).

doi: 10.58286/34001

36. Guo Z, Li Z, Zeng K, Lu X, Ye J, Wang Z. Asymmetric hierarchical acoustic absorber at exceptional point: a loss-induced eigenvalues degeneracy system. *Phys Rev B*. 2024;109(10):104113.

doi: 10.1103/PhysRevB.109.104113

37. Guo Z, Du L, Yang Z, *et al.* Switchable bidirectional sound

absorption via exceptional point modulation in acoustic metastructures with interleaved resonator coupling. *Adv Sci*. 2025;12(44):e08951.

doi: 10.1002/advs.202508951

38. Guo Z, Lei Z, Zeng K, *et al.* Non-integer-dimensional architected materials enabling synergistic acoustic, mechanical, and fluid coupling. *Mater Horiz*. 2026;13(2):748-762.

doi: 10.1039/D5MH01768H

39. Li Z, Wang X, Wang Z, *et al.* Emerging acousto-mechanical metamaterials: from physics-guided design to coupling-driven performance. *Mater Today*. 2025;89:151-171.

doi: 10.1016/j.mattod.2025.06.029

40. Singh SD, Singh MD, Singh AK, Kumar S. AI-enhanced and other load modelling in modern power systems: a comprehensive review of advances, challenges, and future directions. *Rev Int Metod Numer Calc Dis Ing*. 2025;41(4):19.

doi: 10.23967/j.rimni.2025.10.71136

41. Wang Z, Zheng C, Wang L, Gu J, Jing L, Lai X. Optimal design of composite grid/skin structures based on deep learning and Double-Double layup strategy. *Aerosp Sci Technol*. 2024;147:109030.

doi: 10.1016/j.ast.2024.109030

42. Peng J, Ye R, Lan S, Xiao T, Xu C, Song Y. Detection of small rod-end joint bearings via deep feature fusion and confidence propagation clustering. *Rev Int Metod Numer Calc Dis Ing*. 2024;40(3):6.

doi: 10.23967/j.rimni.2024.10.56511

43. Masurkar F, Cui F. A fully optical laser based system for damage detection and localization in rail tracks using ultrasonic Rayleigh waves: a numerical and experimental study. In: Proceedings of the 9th International Conference of the Asian Society for Precision Engineering and Nanotechnology (ASPEN 2022). Singapore: Research Publishing; 2022:717-722.

doi: 10.3850/978-981-18-6021-8_OR-15-0042.html

44. Khalil A, Masurkar F, Abdul-Ameer A. Estimating the reliability of the inspection system employed for detecting defects in rail track using ultrasonic guided waves. In: Al Marri K, Mir FA, David SA, Al-Emran M, eds. BUiD Doctoral Research Conference 2023: Multidisciplinary Studies. Vol 473. Cham: Springer; 2024:190-202.

doi: 10.1007/978-3-031-56121-4_19

45. Masurkar F, Ng KM, Tse PW, Yelve NP. Interrogating the health condition of rails using the narrowband Rayleigh waves emitted by an innovative design of non-contact laser transduction system. *Struct Health Monit*. 2021;20(5):2678-2693.

doi: 10.1177/1475921720967600

46. Masurkar F, Rostami J, Tse PW. Design of an innovative and self-adaptive-smart algorithm to investigate the structural integrity of a rail track using Rayleigh waves emitted and sensed by a fully non-contact laser transduction system. *Appl Acoust*. 2020;166:107354.

doi: 10.1016/j.apacoust.2020.107354

47. Ng KM, Masurkar F, Tse PW, Yelve NP. Design of a new optical system to generate narrowband guided waves with an application for evaluating the health status of rail material. *Opt Lett*. 2019;44(23):5695-5698.

doi: 10.1364/OL.44.005695