A multi-level prognostic framework for delamination-induced failure under compressive fatigue

Ferda C. Gül , Morteza Moradi *, Dimitrios Zarouchas

Center of Excellence in Artificial Intelligence for Structures, Prognostics & Health Management, Aerospace Engineering Faculty, Delft University of Technology, Kluyverweg 1, Delft, 2629 HS, The Netherlands

ARTICLE INFO

ABSTRACT

Predicting the remaining useful life (RUL) of composite structures is particularly challenging in impact-damaged carbon fiber–reinforced polymers (CFRPs) under compressive fatigue, where delamination growth with complex morphology and stochastic progression often governs failure. Guided wave–based structural health monitoring (GW-SHM) enables sensitive damage characterization, yet RUL prediction remains difficult due to the strong dependence of GW–delamination interactions on excitation frequency and damage geometry. Physics-based models often struggle to generalize beyond specific configurations, whereas purely data-driven models can capture complex patterns but typically lack consistency with the underlying physical mechanisms. This study introduces a multi-level, frequency-aware prognostic framework that combines the adaptability of deep learning with the physical interpretability of engineered features. GW signals acquired at multiple excitation frequencies are transformed into time- and time–frequency representations, while damage indicators are derived through temporal segmentation. These indicators are correlated with delamination growth measured by C-scan inspections, providing a link between signal-derived features and physical damage evolution. The multi-level architecture integrates convolutional neural networks, multilayer perceptrons, and long short-term memory layers to capture complementary aspects of degradation. Experimental assessment on seven specimens demonstrates that the proposed framework achieves a minimum mean absolute percentage error (MAPE) of 1.904, corresponding to 11% and 55% improvements over the highest- and lowest-performing single-frequency baselines at 160 kHz and 100 kHz, respectively. The results confirm that integrating GW signal processing with deep learning yields robust and physically consistent RUL predictions for impact-damaged CFRPs, while enhancing the interpretability of prognostic outcomes.

Abbreviations

AI	Artificial Intelligence
CAI	Compression After Impact
CFRP	Carbon Fiber Reinforced Polymer
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
C-C	Compression-Compression
DI	Damage Indicator
DL	Deep Learning
DNN	Deep Neural Networks
EoL	End of Life
GW	Guided Wave
HT	Hilbert Transform
LOOCV	Leave one out cross validation
LSTM	Long-Short Term Memory
MAPE	Mean Absolute Percentage Error
MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
ANN	Artificial Neural Network
PZT	Piezoelectric Transducer
RUL	Remaining Useful Life
ReLU	Rectified Linear Unit
SHM	Structural Health Monitoring
SGD	Stochastic Gradient Descent
WAP	Windowed Averaged Power
AdaBoost	Adaptive Boosting
FCL	Fully Connected Layer

RF   Random Forest
SVR   Support Vector Machine
WAE   Weighted Average Ensemble

1. Introduction

RUL prognostics, a key component of SHM, aims to estimate the time interval before a structure can no longer perform its intended function safely. For composite structures, RUL estimation is particularly complex. These structures, valued for their high strength-to-weight ratio and fatigue resistance, are prone to damage types that differ fundamentally from those in metals. Among them, impact-induced damage, such as that caused by bird strikes or dropped tools, can lead to delamination, matrix cracking, and fiber–matrix debonding, all evolving in complex and often unpredictable ways [1,2]. Delamination is especially critical under compressive loading, where it may initiate at multiple interfaces, propagate unstably, and severely compromise structural integrity [3,4].

Maintenance strategies for composite structures typically follow safe-life or damage tolerance principles, with regulatory standards requiring periodic inspection and damage assessment to prevent catastrophic failure [5]. However, composites' anisotropic and heterogeneous nature makes it particularly challenging to quantify damage severity and predict time to failure reliably [6,7]. In metallic structures, physics-based approaches such as fracture mechanics and Paris' law have been successfully employed for RUL prediction, since crack growth under cyclic loading can often be described by relatively simple and continuous propagation laws [8,9]. These assumptions, however, do not transfer well to composites, where degradation involves interacting, anisotropic, and often stochastic failure modes [10]. To overcome these limitations, data-driven methodologies have emerged as alternative pathways for RUL prediction. Such methods generally fall into two categories: direct prediction and damage-indicator (DI)-based approaches [11]. Direct models estimate RUL from raw or minimally processed sensor data, without explicitly modeling the physical degradation process. In contrast, DI-based models use signal-derived features to represent the structure's health or damage state. These indicators can then be mapped to RUL using statistical or machine learning techniques, providing a more physically meaningful link between observed signals and degradation behavior [12,13].

Among sensing techniques employed for composite structures, GW-SHM stands out due to its powerful interrogating capability and high sensitivity to internal damage [14,15]. Particularly well-suited for aerospace applications, GW-SHM enables the detection of barely visible or subsurface damage, such as delamination, with minimal sensor deployment and over large structural areas [16,17]. GW interrogation offers a sensitive and continuous means of assessing damage severity, thereby providing a reliable basis for RUL prediction, enabling both direct and DI-based prognosis [17]. GW's effectiveness has been demonstrated in tasks such as delamination detection, localization, and severity assessment in composite components. including fuselage sections and wing panels [18–21]. Several studies have explored GW-based approaches for monitoring damage propagation; their extension toward RUL prognostics, particularly under compressive fatigue conditions, remains limited. Traditional state estimation techniques, such as Kalman and particle filters, have been applied to model degradation using a small set of GW features (e.g., amplitude, energy, arrival time) [22–24]. Mishra et al. combined principal component regression with a Wiener process to predict delamination growth from GWs, but by assuming a single stochastic growth mode, their approach remains limited for impact-induced delamination and realistic RUL prognosis in composites [25]. However, these approaches often rely on assumptions of linearity, Gaussian noise, and stationary system dynamics, which struggle to capture the nonlinear and unstable progression of delamination driven by compressive fatigue loading [26]. In addition, their reliance on single-frequency inputs limits the ability to exploit the frequency-dependent nature of wave–damage interactions. As a result,

these filtering-based methods struggle to generalize across structural geometries and loading regimes, reducing their practical applicability in high-reliability engineering contexts.

Recent efforts have turned to deep learning (DL) based data-driven methods DL, which are particularly well-suited for GW data analysis, as they can automatically extract complex patterns and latent features without manual intervention [27–30]. Architectures such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks have shown strong performance in learning spatial and temporal degradation features, respectively, making them ideal candidates for RUL prediction tasks [31–34]. Feng et al. investigated low-velocity impact localization in CFRP plates using GW time–frequency features combined with CNNs [35]. Moradi et al. designed history-independent health indicators for composite structures by extracting multi-level GW characteristics and using CNNs as base learners under a semi-supervised framework; however, a limitation of this approach is the lack of linkage to physical damage states, reducing correlation with actual failure mechanisms [36]. DL models also support advanced data fusion strategies, enabling the integration of multi-level features from different SHM modalities and sensor configurations [37,38].

Based on the reviewed literature, there is a need for more advanced GW-SHM approaches to support reliable RUL prognostics in composites. Existing statistical models frequently rely on simplified assumptions that are insufficient to capture the nonlinear and unstable progression of delamination under compressive fatigue. Considering the complexity of delamination and its interaction with GWs, it is notable that despite the proven interrogating capabilities of GWs for composite damage detection, their application to RUL prognostics remains limited in the literature. The prevalence of single-frequency analyses further highlights the need for multi-frequency strategies to exploit the frequency-dependent sensitivity of GW–damage interactions. Finally, although data-driven methods have demonstrated strong predictive performance, there remains a need to reinforce their linkage to physical damage states to enhance interpretability in predictions and ensure applicability in safety-critical RUL prognosis.

To address the limitations of current RUL prediction methods in composite structures, this study introduces a novel, frequency-aware, multi-level DL framework that emphasizes accuracy and RUL interpretability. The proposed framework leverages the full informational richness of GW signals obtained under multiple excitation frequencies. Each excitation frequency is treated as an independent source of degradation information. From GW signals, both raw representations (e. g., Hilbert envelope and wavelet-transformed signals) and damage indicators (GW-DIs) are extracted to characterize delamination progression. These heterogeneous inputs are processed through a multi-level learning architecture consisting of two main stages: base and ensemble learners, and fusion models. Accordingly, this study makes three key contributions:

- A frequency-aware, multi-level prognostic framework that integrates GW signal processing with DL, enabling robust and accurate RUL prediction for CFRPs under compressive fatigue.
- An investigation of feature–damage correlations showing that GW-DIs, even when only weakly correlated with delamination measured by C-scan, can still contribute substantially to RUL prediction.
- An experimental assessment conducted on seven impact-damaged CFRP specimens under compression–compression (C–C) fatigue, in which the framework's prognostic performance is evaluated using leave-one-out cross-validation (LOOCV), achieving improvements of up to 55% compared to single-frequency baselines.

The proposed methodology is applied to the dataset comprising GW signals collected from a series of C-C fatigue experiments on seven woven-type CFRP specimens [39]. Initial damage was introduced through controlled low-velocity impacts, and specimens were subjected
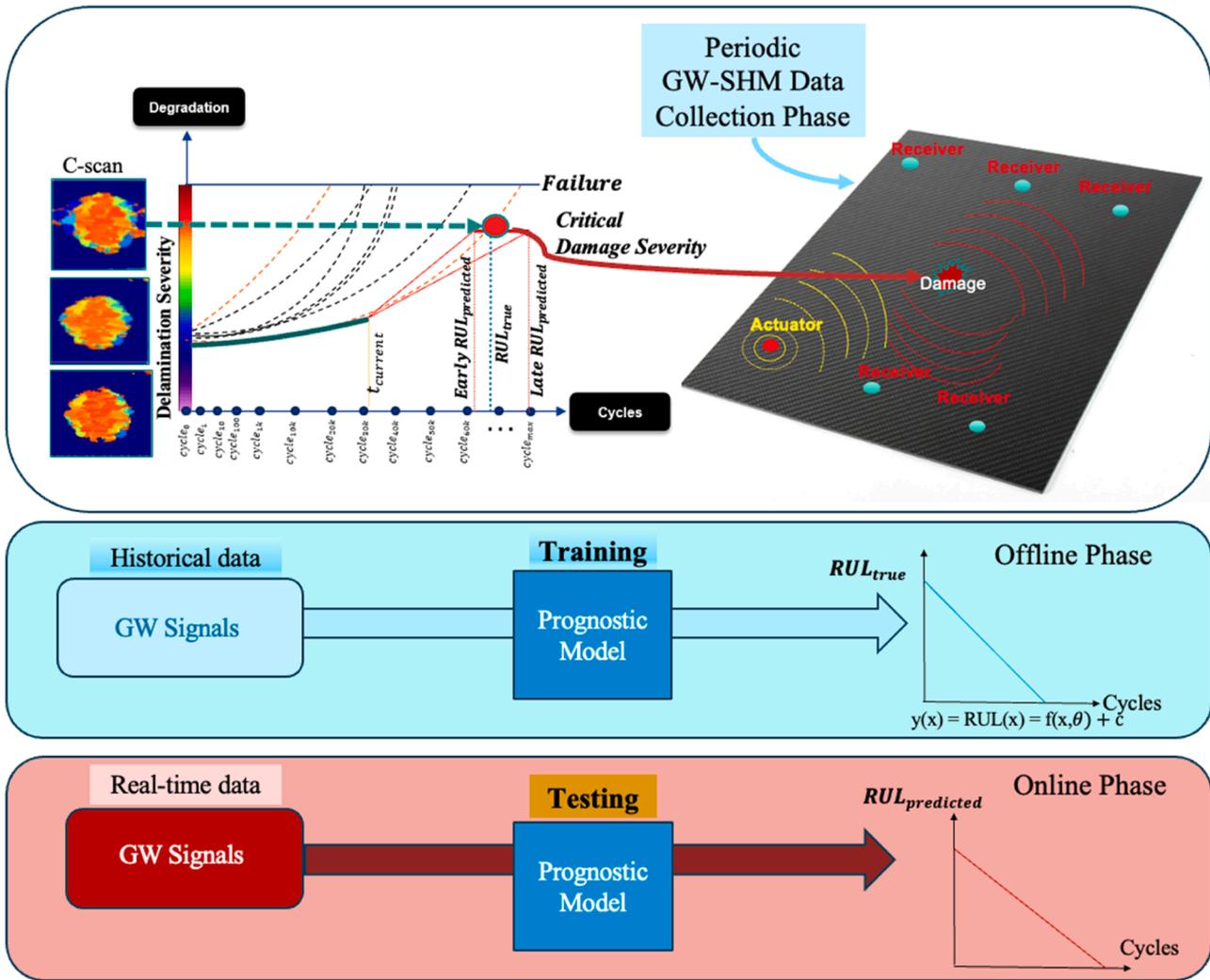
Fig. 1. Schematic illustration of the proposed GW-SHM based prognostic framework.

to compressive fatigue loading until failure. This setup yielded a diverse dataset with varying delamination growth trajectories, end-of-life conditions, and loading histories. The proposed model performance was evaluated using LOOCV. In each fold, the model was tested on data from a specimen excluded from training, allowing a realistic assessment of generalization capability under unseen structural scenarios. In addition to the proposed DL framework, three classical machine learning regressors, including support vector regression (SVR), random forest (RF), and adaptive boosting (AdaBoost), were implemented as benchmarks. SVR was selected for its robustness in small-sample, nonlinear regression settings; RF provides strong generalization through variance reduction and insensitivity to feature scaling; and AdaBoost offers an error-focused boosting mechanism that prioritizes difficult-to-predict samples [40, 41]. Together, these models offer complementary strategies to benchmark the performance of the proposed approach.

The structure of this paper is as follows. Section 2 provides the theoretical and conceptual background relevant to this study, introducing the concept of RUL in the context of GW-SHM and GW-damage interaction principles. Section 3 outlines the proposed prognostic methodology, presenting the multi-level DL framework. Section 4 describes the experimental procedures, including the low-velocity impact and C-C fatigue testing, as well as the data acquisition process. Section 5 presents and discusses the results, providing a detailed performance evaluation of the proposed framework and comparing the outcomes across different base learners and fusion strategies. Finally, Section 6

concludes the paper by summarizing the contributions and outlining potential paths for future research.

2. Motivation & background

2.1. Delamination-induced failure prognosis

Damage tolerance requires that structures sustain load despite damage, ensuring sufficient time for inspection and repair. According to EASA AMC [5], early detection and timely repair are essential to preserve these characteristics. In composites, however, heterogeneous microstructures and irregular degradation patterns complicate the definition of critical damage size, and the prediction of RUL. Impact-induced delamination can bypass early stages and propagate rapidly under cyclic loading, shortening fatigue life [42]. Delamination progresses gradually in Mode I under tensile loading, while compressive loading triggers buckling-induced Mode II/III mechanisms that accelerate growth. Furthermore, in compression-after-impact scenarios, delamination growth is driven by the out-of-plane buckling of surface plies, which alters the local stress distribution and accelerates delamination progression. Once the structure reaches a critical damage state or stiffness threshold, localized buckling of surface plies further accelerates degradation, significantly reducing the component's structural integrity [43–46].

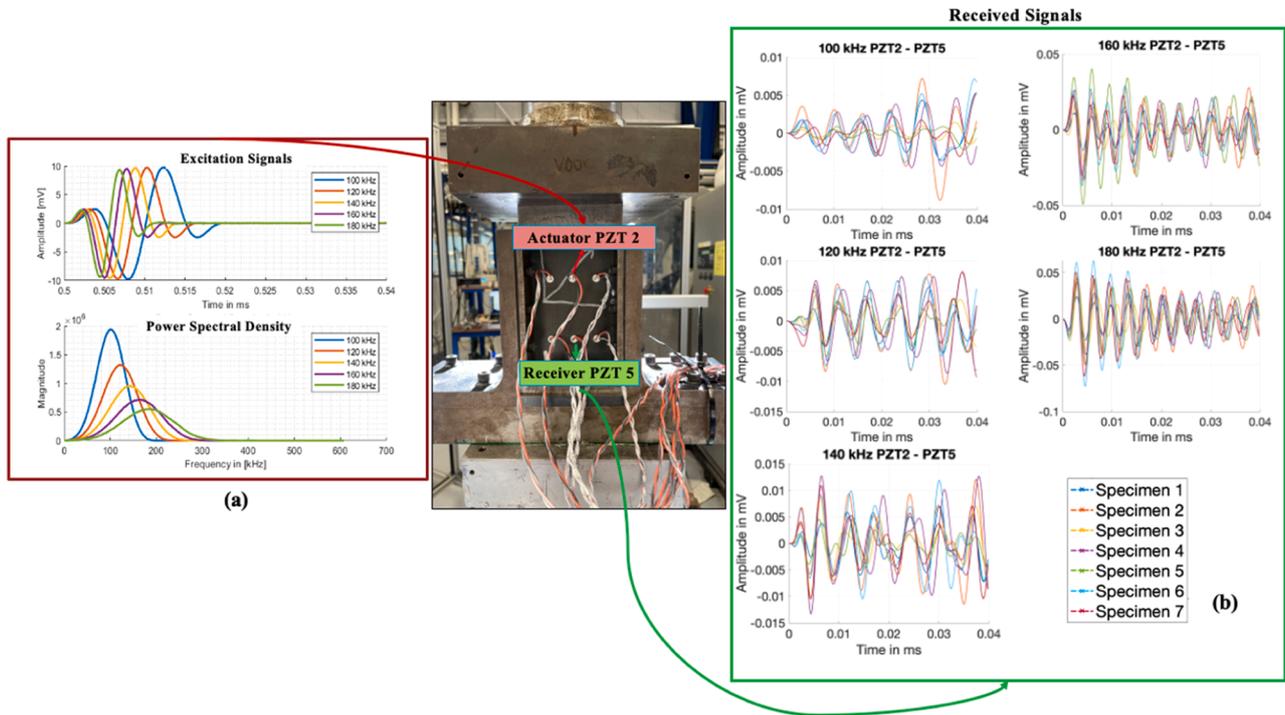Delamination evolution's complexity and its propagation's

Fig. 2. (a) Excitation signals at different central frequencies f_{exc} and their corresponding power spectral density. (b) Received signals at PZT5 for excitations from actuator PZT2 at 100–180 kHz at healthy state, obtained from seven different specimens. The central image shows the sensor installed specimen.

unpredictable nature under compressive fatigue conditions require further research to enhance failure prognostics methodologies. The schematic representation in Fig. 1 is therefore not intended to reproduce the underlying fracture mechanics, but rather to illustrate the conceptual basis of RUL estimation from a maintenance-oriented perspective. This framework employs delamination size as a measurable degradation indicator, reflecting that inspection and maintenance actions are ultimately governed by observable damage severity. In the present study, C-scan measurements of delamination length are used for labeling purposes, enabling the GW-DIs to be quantitatively correlated with physical damage states [47]. This correlation allows to identify the contribution of different knowledge-based GW-DIs to RUL prediction, and their performance can be critically discussed together with direct RUL predictions. In Fig. 1, schematic illustration of the proposed GW-SHM based prognostic framework is shown. In the figure, the $RUL_{threshold}$ represents the point at which the degradation accelerates and followed by failure. The final prediction of the RUL model incorporates the interval between critical damage severity and failure. Critical damage severity corresponds to the time that the structure reaches a specific delamination length that defines its final delamination state. This delamination length is expected to vary across tested samples, not only due to differences in loading conditions and initial damage characteristics, but also as a natural consequence of the stochastic nature of delamination evolution in CFRP laminates.

2.2. SHM data collection: GW signals

GWs are elastic waves in plate-like structures, propagating as coupled longitudinal and shear modes. Their behavior is governed by structural geometry, ply orientation, excitation angle, and frequency. Progressive damage alters material properties through mode conversion, complicating frequency-dependent modeling and strongly influencing propagation mechanisms. At lower excitation frequencies, the antisymmetric A0 mode dominates, with longer wavelengths suited to global stiffness changes and large-scale degradation. Higher excitation frequencies more effectively activate the symmetric S0 mode, with

shorter wavelengths and higher dispersion, making it sensitive to localized damage such as delamination or matrix cracking [48]. In practice, however, received signals contain a superposition of A0, S0, and higher-order modes, due to both damage-induced mode conversion and spatial sensor–actuator effects, leading to overlap in time and frequency domains and complicating interpretation [49].

Employing a range of excitation frequencies thus provides a more comprehensive basis for GW-based RUL prognostics, as it leverages the complementary damage sensitivities of different modes while reducing reliance on mode-specific assumptions [50]. In the experiment, GW excitation signals are performed as 2-cycle tone-burst signals at the center frequency f_{exc} of 100 kHz, 120 kHz, 140 kHz, 160 kHz, and 180 kHz. The equation for the excitation signal is given as following:

$$x(t) = \frac{A}{2}\left[1 - \cos\left(\frac{2\pi f_{exc}}{N_{cycle}}t\right)\right]\sin(2\pi f_{exc}) \tag{1}$$

where A = 10 V, N_{cycle} = 2. Excitation signal is presented in Fig. 2 showing the signal in time and frequency domain for each center frequency f_{exc} of excitation signal with power spectral density which describes how the power of a signal is distributed across different frequencies, representing the signal's power per unit frequency (e.g., W/Hz) and is computed as the squared magnitude of the Fourier Transform normalized by signal duration.

3. Prognosis methodology

3.1. General framework

The proposed framework performs RUL estimation through a hierarchical architecture composed of three stages: base learners (Level 0), an ensemble learner (Level 1), and a multi-frequency fusion model (Level 2). This design allows the model to exploit both single-frequency and cross-frequency information extracted from GW signals. Benchmark models are also used for global performance comparison. The three levels are further explained in the following:
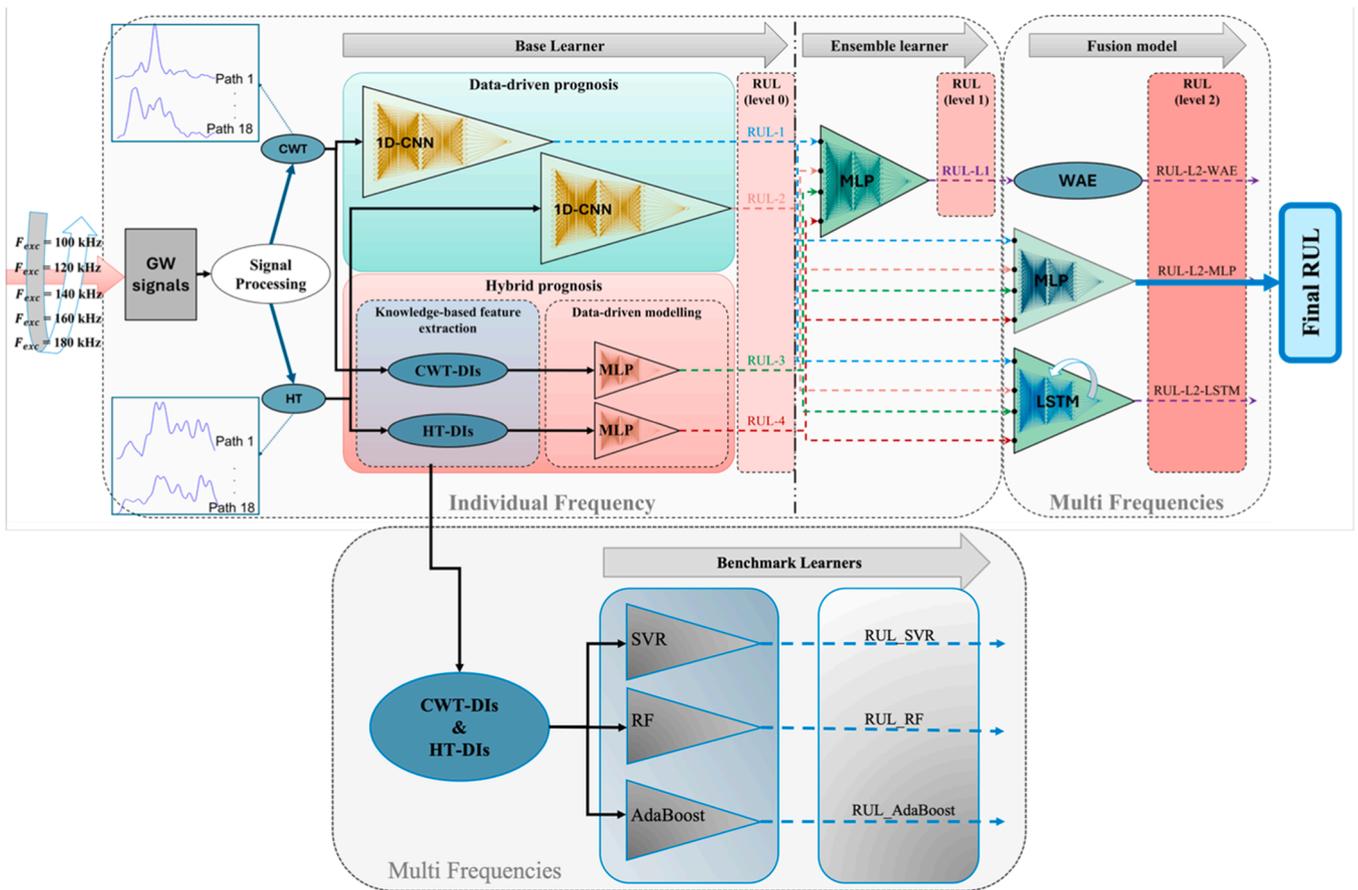
Fig. 3. Overall proposed framework including base learners and fusion steps at different levels.
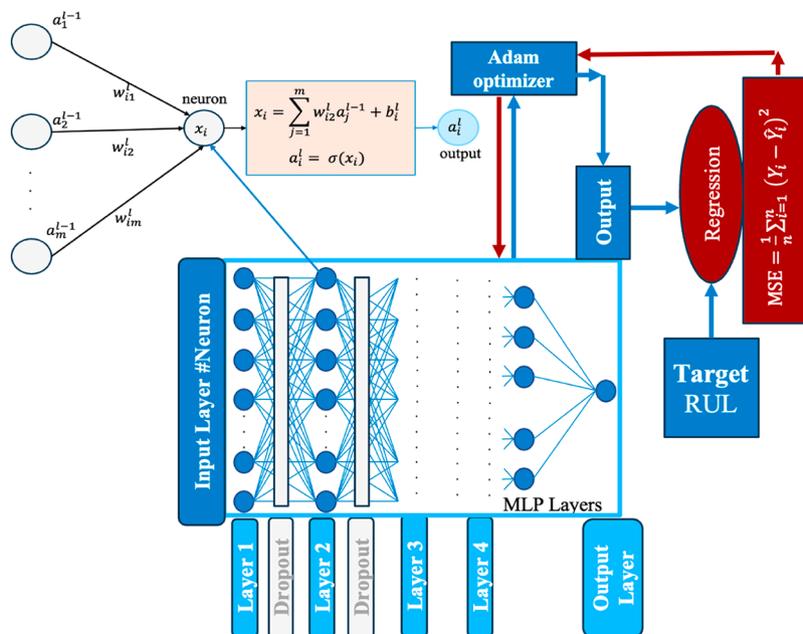

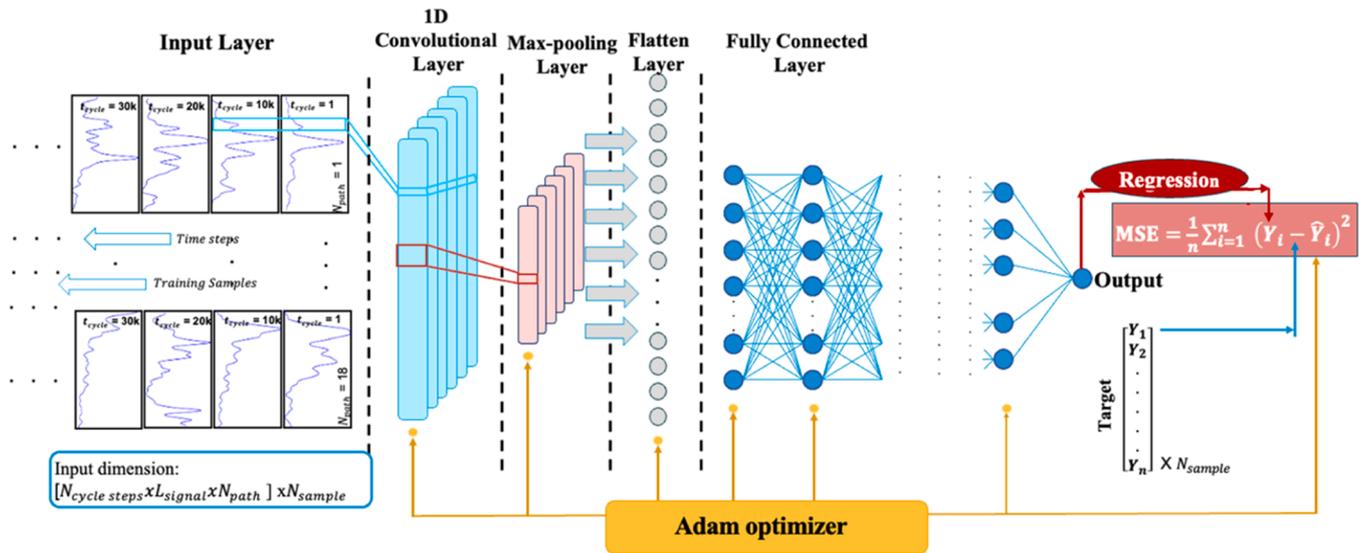
Fig. 4. MLP architecture for base and fusion learners.

Fig. 5. 1D-CNN base learner.
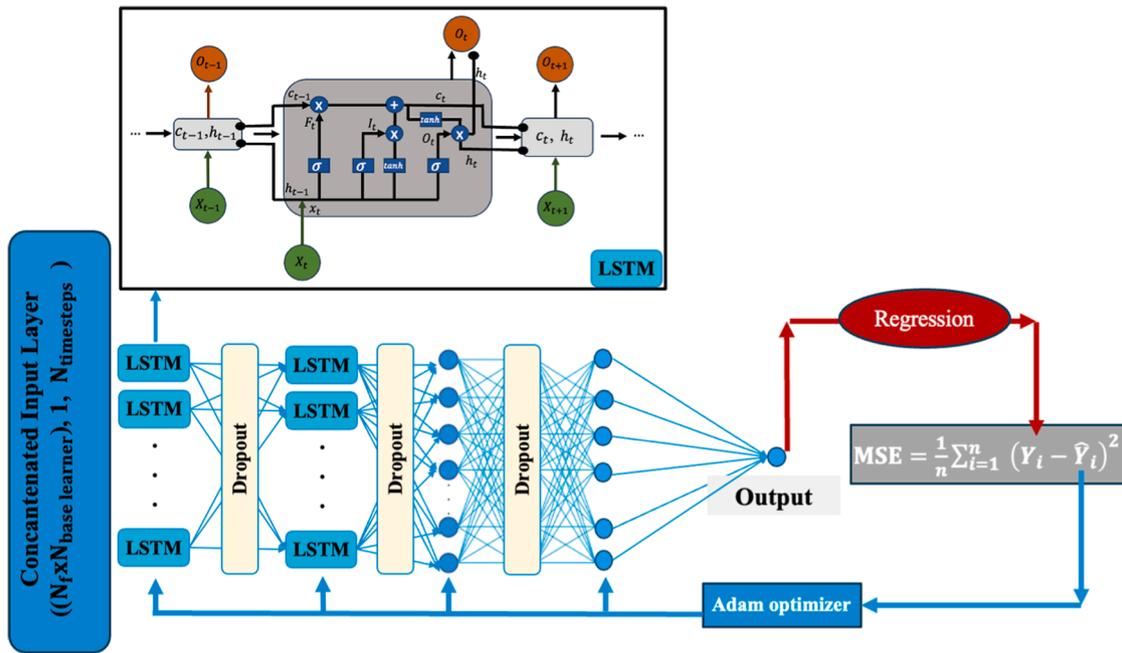
Fig. 6. LSTM model.

- **Level 0 – Base learner (individual frequency models):**

At the base learner level, two complementary RUL prediction strategies are implemented: First, in the data-driven prognosis approach, GW signals are transformed using time–frequency representations (CWT) and time-domain processing (Hilbert Transform), after which frequency-specific 1D-CNNs are trained to map these inputs directly to RUL. Second, in the hybrid prognosis approach, knowledge-driven damage indicators (GW-DIs) extracted from the CWT and HT domains are used as structured features and fed into MLP regressors to produce RUL estimates. Consequently, for every excitation frequency and each input modality (CWT, HT, CWT-DIs, HT-DIs), a separate base learner (either 1D-CNN or MLP) is trained, allowing the framework to capture complementary characteristics of degradation and yield diverse predictive behaviors.

- **Level 1 – Ensemble learner:**

The predictions of the base learners at each frequency are then combined through an MLP-based ensemble learner. This step integrates heterogeneous information sources, raw signals, and engineered indicators, allowing the system to leverage both deep and knowledge-driven feature extraction. The ensemble produces a unified RUL estimate for each GW excitation frequency (i.e., RUL-L1). This level plays a key role in reducing modality-specific bias and improving stability by aggregating complementary prediction patterns to assess the performance of each excitation frequency based on information in the RUL output.

- **Level 2 – Fusion model (multi-frequency integration):**

The fusion-level RUL estimates are further fused through three

Fig. 7. (a) Large healthy state panel; (b) cutting layout.
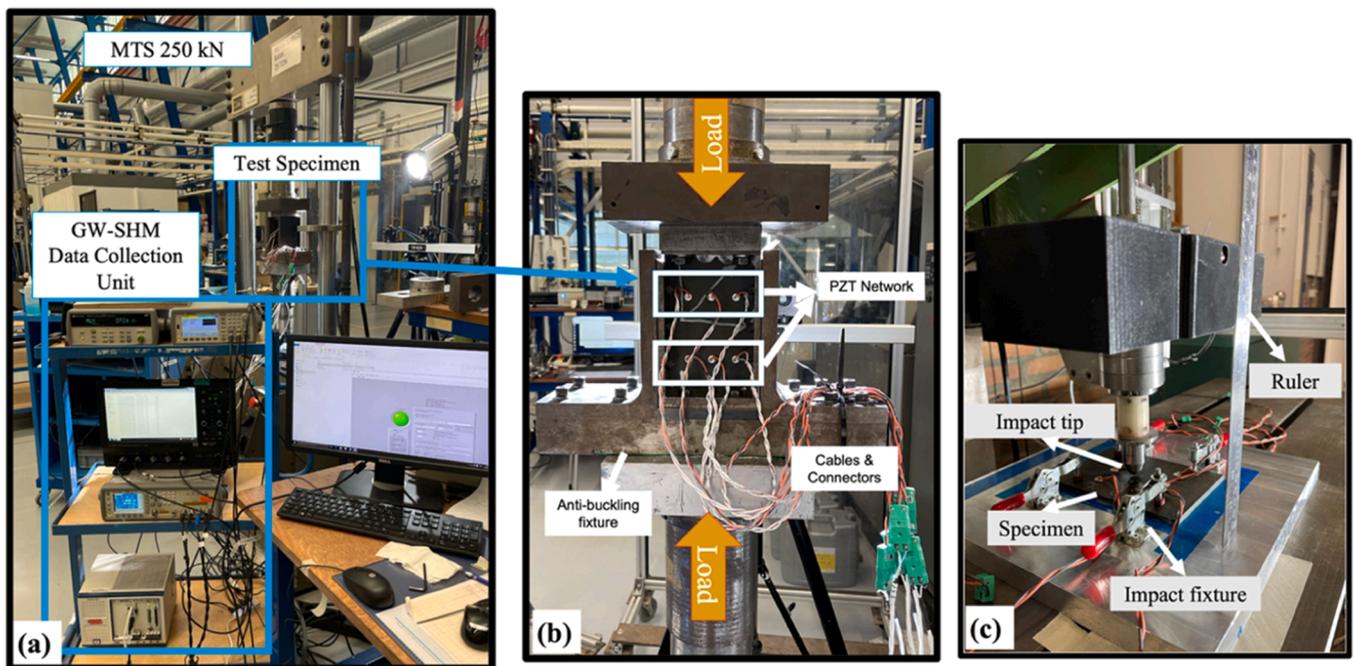
Fig. 8. (a) SHM unit and fatigue testing MTS machine (b) Anti-buckling fixture and sensor-installed sample (c) Impact testing setup.
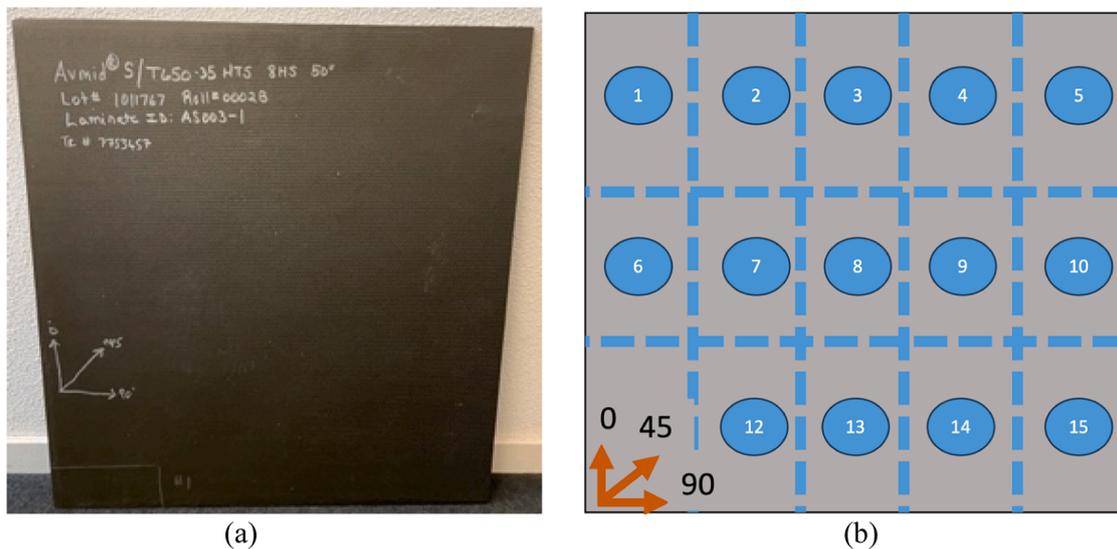
strategies, allowing the framework to exploit the inherently multi-frequency nature of GW inspection and enhancing the robustness of the final RUL estimate. In the weighted average ensemble (WAE), frequency-specific RUL outputs are combined through optimally tuned weights to produce a global estimate. The MLP fusion strategy captures nonlinear relationships across frequencies, allowing richer cross-frequency interactions. Meanwhile, the LSTM fusion approach treats the frequency-wise predictions as an ordered sequence, modeling implicit dependencies among frequencies and providing an additional pathway for improving the stability and accuracy of the final RUL estimate.

To evaluate the performance of the proposed framework, three classical ML regression algorithms, including SVR, RF, and AdaBoost, are trained using multi-frequency GW-DIs (CWT-DIs and HT-DIs). These models operate independently of the DL pipeline and provide baseline RUL estimates, named RUL_SVR, RUL_RF, and RUL_AdaBoost.

Benchmark learners highlight the added value of the proposed multi-level DL framework relative to established ML approaches.

The final model choice is made empirically through comparative evaluation across all candidate fusion strategies. Since the MLP-based fusion achieved the lowest prediction error and the most stable generalization across test scenarios, it is selected as the final fusion step of the proposed deep learning approach. Fig. 3 illustrates the full hierarchical structure of the framework, showing base learners operating at individual frequencies, ensemble learners combining multiple modalities, and fusion models integrating all excitation frequencies into the final RUL output.

3.2. Signal processing for GWs

There are some well-adapted signal processing techniques that enable the analysis of GW signals to convert them into DIs that can be

Fig. 9. PZT network and sensor placement.

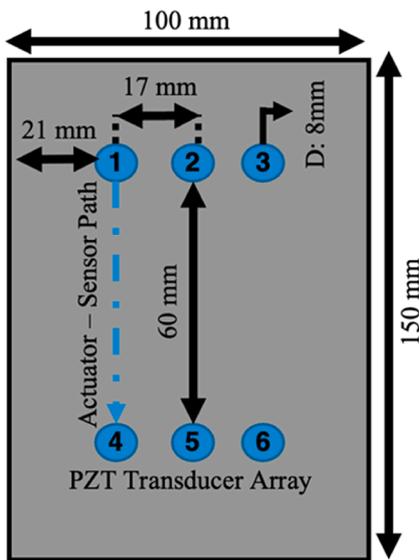further organized as prognostic features [51]. In the time domain, a windowing strategy can be applied to isolate the first-arrival wave packet, which carries direct information about the propagation path

prior to significant interactions with damage. For the subsequent portion of the signal, which is more likely to include reflections and scattered components induced by discontinuities, time–frequency representations can be employed to construct localized time windows. In this work Hilbert Transform (HT) and Continuous Wavelet Transform (CWT) are used to obtain DIs from GW signals. HT is implemented to obtain signal envelopes from residual GW signal that allows for the calculation of the signal energy in desired time interval and DI-Env are obtained with Eqs. (2)–(4):

$$H(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} \mathrm{d}\tau \tag{2}$$

$$z(t) = x(t) + iH(t) \tag{3}$$

$$DI - Env = \frac{\left[\sum_{t}^{T_{window}} H(t)\right]_n}{\left[\sum_{t}^{T_{window}} H(t)\right]_{n=1}}, \ n = 1, \dots N_{cycle\ steps} \tag{4}$$

A time window was assigned with different lengths for GW signals actuated at different f_{exc}, resulting in a longer time window for a lower frequency of 100 kHz and a shorter one for 180 kHz. These intervals are [0-5], [0-4.5], [0-4], [0-3.5], and [0-3] μs. Given that GWs are non-stationary and time-varying signals, time-frequency analysis such as the CWT provides a suitable representation for further feature extraction and is widely used methods in the literature [52,53]. CWT is employed
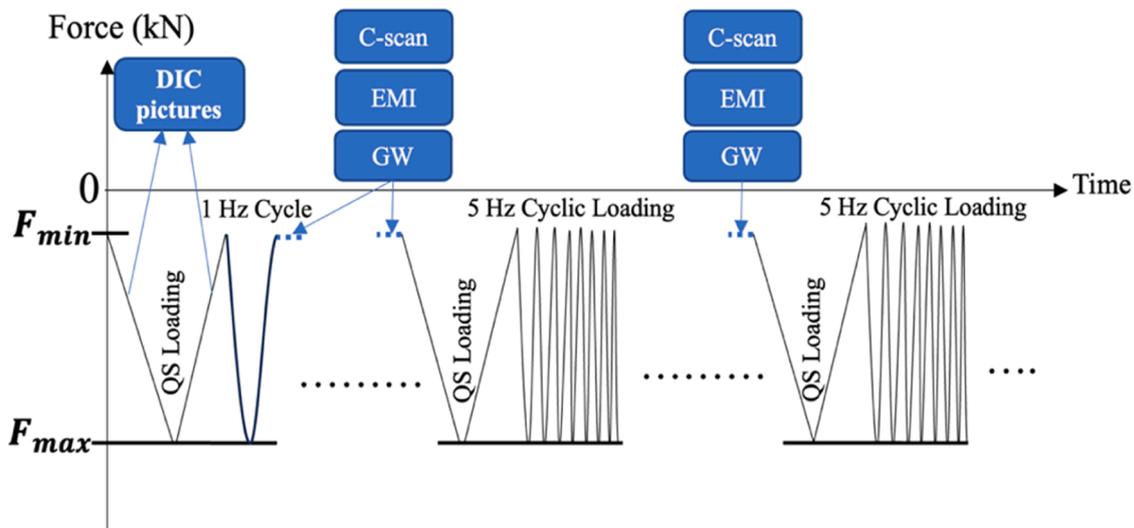
Fig. 10. Fatigue loading and data acquisition steps.

Table 1
Information of composite specimens tested under run-to-failure compression-fatigue loading after impact.

#No of specimens	Impact energy (joule)	Max/Min force (- kN/- kN)	EoL (fatigue cycles) x1000	Time to max measured delamination level x1000	Number of measurement (Time) Steps	Max applied force (-kN)
1	19.45	135 / 13.5	79	70	8	140
2	15.35	140 / 14	40.5	40	5	140
3	15.35	140 / 14	97.2	90	10	140
4	19.45	130 / 13	38.5	35	4	137
5	16.31	140 / 14	27.9	25	3	140
6	15.35	140 / 14	21.3	20	4	140
7	16.31	135 / 13.5	15.5	15	5	135

Fig. 11. C-scan measurements of impact-induced delamination propagates under C-C fatigue loading.
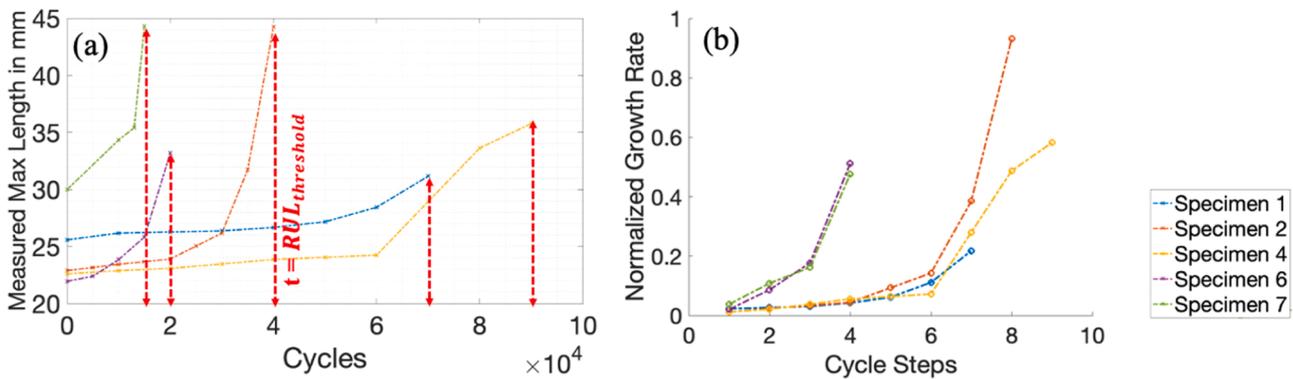
Fig. 12. (a) Maximum delamination length estimated by C-scan during fatigue (b) calculated growth rate.

mainly for GWs as it allows for capturing localized features in the signal, making it suitable for analyzing signals with abrupt changes or transient events. In CWT analyses, wavelets-localized functions or waveforms are utilized to examine different segments of a signal across various scales. The wavelet function is represented by Ψ, and the continuous CWT coefficients are expressed as CWT (a, b). When a signal segment closely matches the form or pattern of the wavelet, the wavelet coefficients reach their maximum value. The Morlet wavelet provides enhanced signal decomposition and analysis, capturing essential features with minimal distortion compared to other wavelets [54]. The t - T_{window} is defined for windowed average power calculation (WAP) based on the peak amplitude of the average power (AP) obtained in the last cycle step, and the total energy variation is quantified between the local minima of the detected peak where the related equations are given in Eqs. (5)–(8):
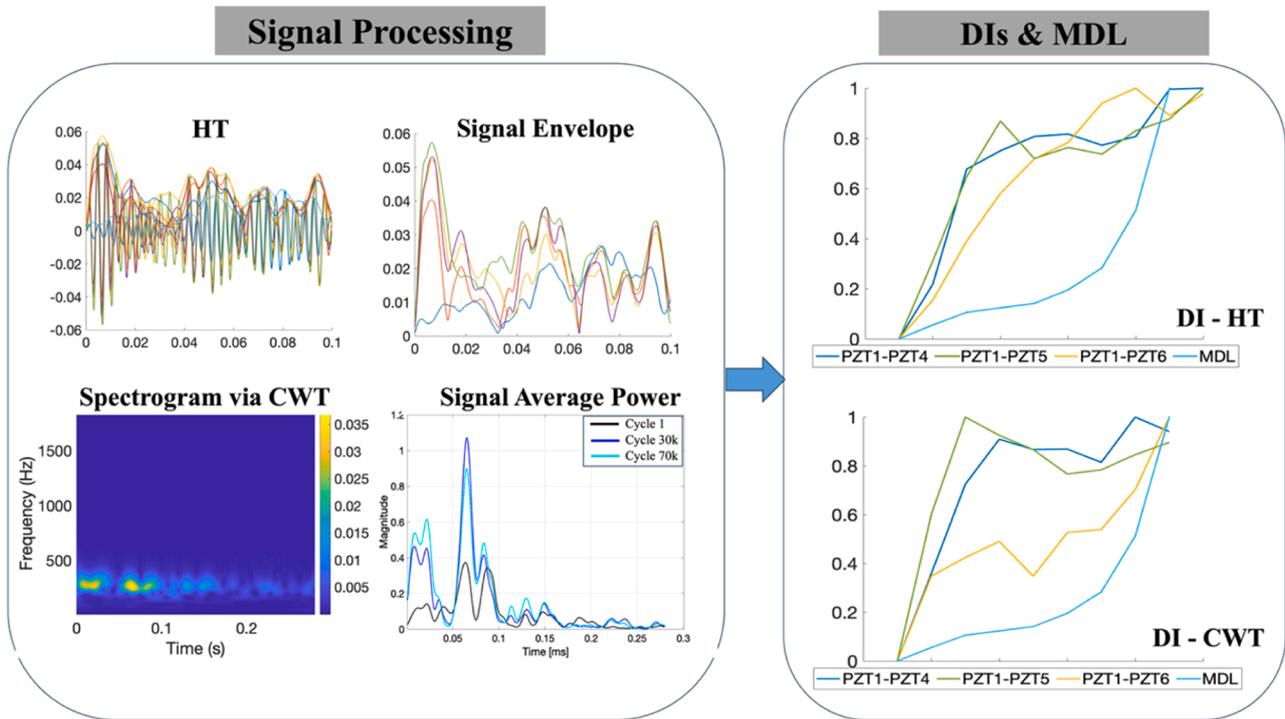
Fig. 13. GW signals for 180 kHz f_{exc} in time and time-frequency domain and GW-DIs for actuator PZT1 sensors PZT4-5-6 for Specimen 1.
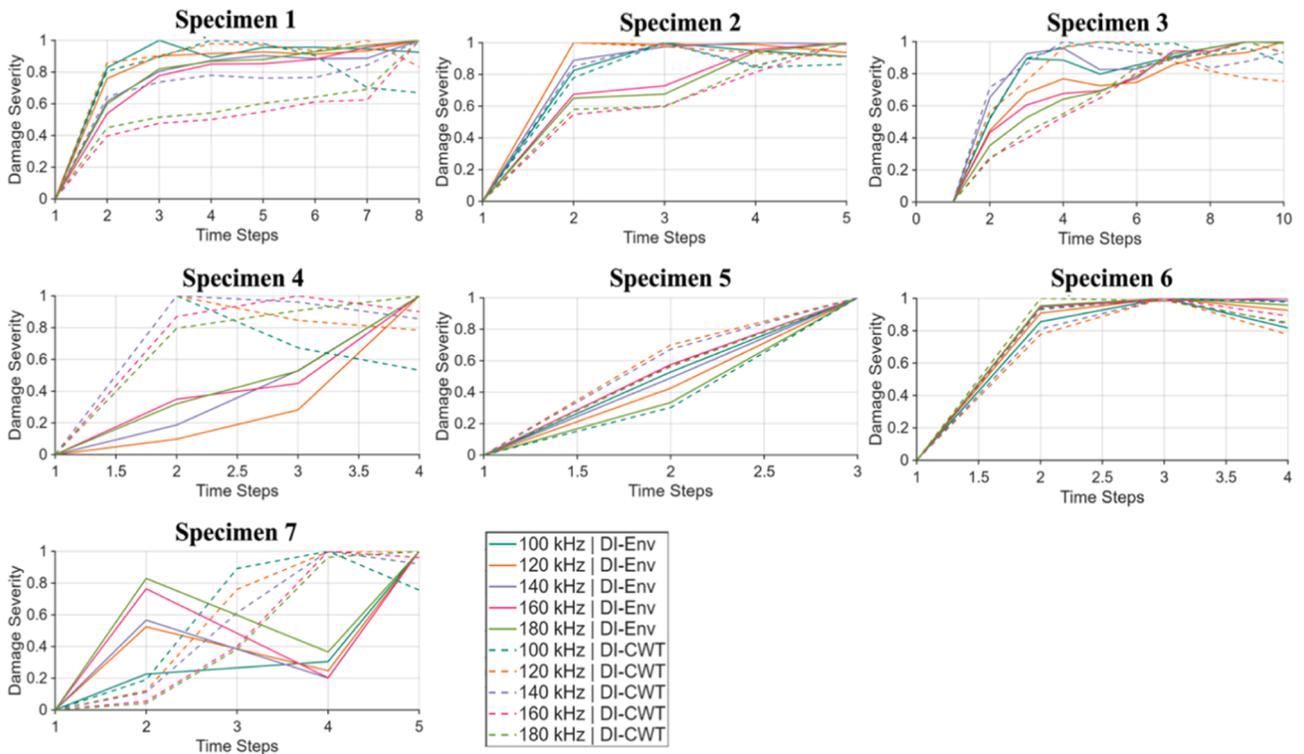
Fig. 14. DI-Env and DI-CWT for all specimens for each f_{exc}.

$$CWT(a,b) = \int \psi\left(\frac{t-b}{a}\right)x(t)dt \qquad (5)$$

$$x(t) = |CWT(a,b)| \qquad (6)$$

$$AP(t) = \frac{\left(\sqrt{\sum_{j=1}^{n} x(t)_j^2}\right)^2}{M} = \frac{\sum_{j=1}^{n} x(t)_j^2}{M} \qquad (7)$$

$$DI - CWT = WAP(t) = \frac{\left[\sum_{t}^{T_{window}} AP(t)\right]_n}{\left[\sum_{t}^{T_{window}} AP(t)\right]_{n=1}} \quad n = 1, \dots N_{cycle\ steps} \qquad (8)$$
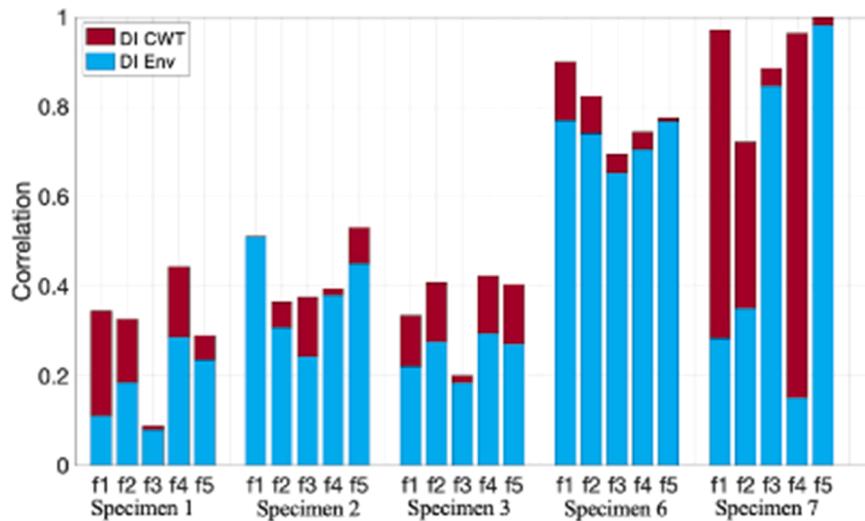
Fig. 15. Path-averaged correlation values for DI-Env and DI-CWT vs measured delamination length.

Additionally, the correlation metric, Corr, is used to evaluate the similarity between DIs and observed damage degradation trends. It quantifies the extent to which the DIs accurately reflect the damage propagation by correlating them with the measured delamination length. A high correlation indicates that a DI present similar behavior as the measured delamination length. The metric Corr, given in Eq. (9), is rated on a scale from 0 to 1, with a score of 1 representing higher score indicates higher similarity.

$$\text{Corr} = \max_{j} \left| \frac{\text{cov}\left(\text{norm}\left(x_j\right), \text{norm}\left(D_j\right)\right)}{\text{std}\left(x_j\right)\text{std}\left(D_j\right)} \right|, j = 1, 2, ..., M \qquad (9)$$

Here, cov denotes the covariance, while $\text{std}(x_j)$ and $\text{std}(D_j)$ represent the standard deviations of the normalized variables x_j and D_j, respectively. The formulation is equivalent to the absolute Pearson correlation coefficient, with the maximum value over M signal paths considered. For each degradation history, GW datasets from individual actuator–sensor paths were standardized via z-score normalization given in Eq. (10):

$$Z = \frac{X - \mu}{\text{std}(X)} \qquad (10)$$

where Z is the standardized value, X is the original data point, μ is the mean, and $\text{std}(X)$ denotes the standard deviation of the GW input vector corresponding to a given sensor–actuator path at a specific excitation frequency.

3.3. Base and ensemble learners

In the initial stage of the proposed framework, base learners are applied separately to each f_{exc} and input type, specifically DI-CWT, DI-Env, CNN-CWT, and CNN-Env, resulting in a total of 20 RUL predictions per test specimen. Following the base learners, the ensemble learner phase, referred to as RUL-L1, concatenates the outputs from the four base learners feeds them into an MLP model. This results in a single aggregated RUL prediction for each f_{exc}, producing five predictions per specimen. In this study, stacking was adopted as the ensemble strategy because alternative methods do not align with the heterogeneous and multi-frequency structure of the proposed framework. Stacking enables a learner to selectively integrate heterogeneous predictors across all excitation frequencies, suppressing unstable or uninformative models while amplifying those that capture the most reliable degradation patterns. This allows the final RUL estimate to be driven primarily by the most informative frequency bands and modalities, making stacking the most robust and conceptually appropriate ensemble approach for the proposed multi-level prognostic framework. On the other hand, alternative ensemble methods have limitations for this problem. Bagging relies on averaging predictions from models trained on bootstrapped samples, which mathematically forces the ensemble toward a mean estimator and prevents it from exploiting complementary feature–model relationships. Boosting sequentially reweights samples based on residual errors, effectively amplifying high-loss points; however, in noisy and stochastic regimes such as GW signals, this leads to instability and propagation of noise into the final predictor [55].

3.3.1. Multilayered perceptron (MLP)

In this study, MLP models are utilized as both base, ensemble and fusion learners, with the learning scheme of the proposed MLP structure illustrated in Fig. 4. In a standard MLP architecture, multiple hidden layers are placed between the input and output layers. The depth of the network, reflected by the number of layers, enables the model to learn abstract representations of the input, which is particularly beneficial in handling high-dimensional and non-linear relationships. Eq. (11) describes the computation of the output from the m-th neuron in layer l, where a_m^l represents the output, $a_m^{(l-1)}$ is the input from the previous layer, w_{jm}^l is the weight between neurons, and b_j^l is the bias term, with the activation function σ applied to the weighted sum.

$$a_m^l = \sigma\left(z_j^l\right) = \sigma\left(\sum_m w_{jm}^l a_m^{l-1} + b_j^l\right) \qquad (11)$$

The activation function, denoted as σ, can be either linear or nonlinear and allows MLP to capture complex data patterns. Well-adapted activation functions include the rectified linear unit (ReLU), sigmoid, and hyperbolic tangent (tanh) given in Eq. (12):

$$\text{ReLU}(x) = \max(0, x), sigmoid(x) = \frac{1}{1 + e^{-x}}, \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \qquad (12)$$

3.3.2. Convolutional neural network (CNN)

The core building block of a CNN is the convolutional layer, which applies a set of filters (or kernels) across the input data to detect local patterns. While CNNs are often associated with 2D data like images, they can also be adapted for 1D data, such as time series [56]. In a 1D-CNN, the convolutional filters slide along one dimension of the data, making them ideal for capturing temporal patterns in sequences. This approach is particularly effective for tasks like time series regression, where the goal is to predict a continuous output based on past values in a sequence. By applying 1D convolutions, the network can learn to detect trends,
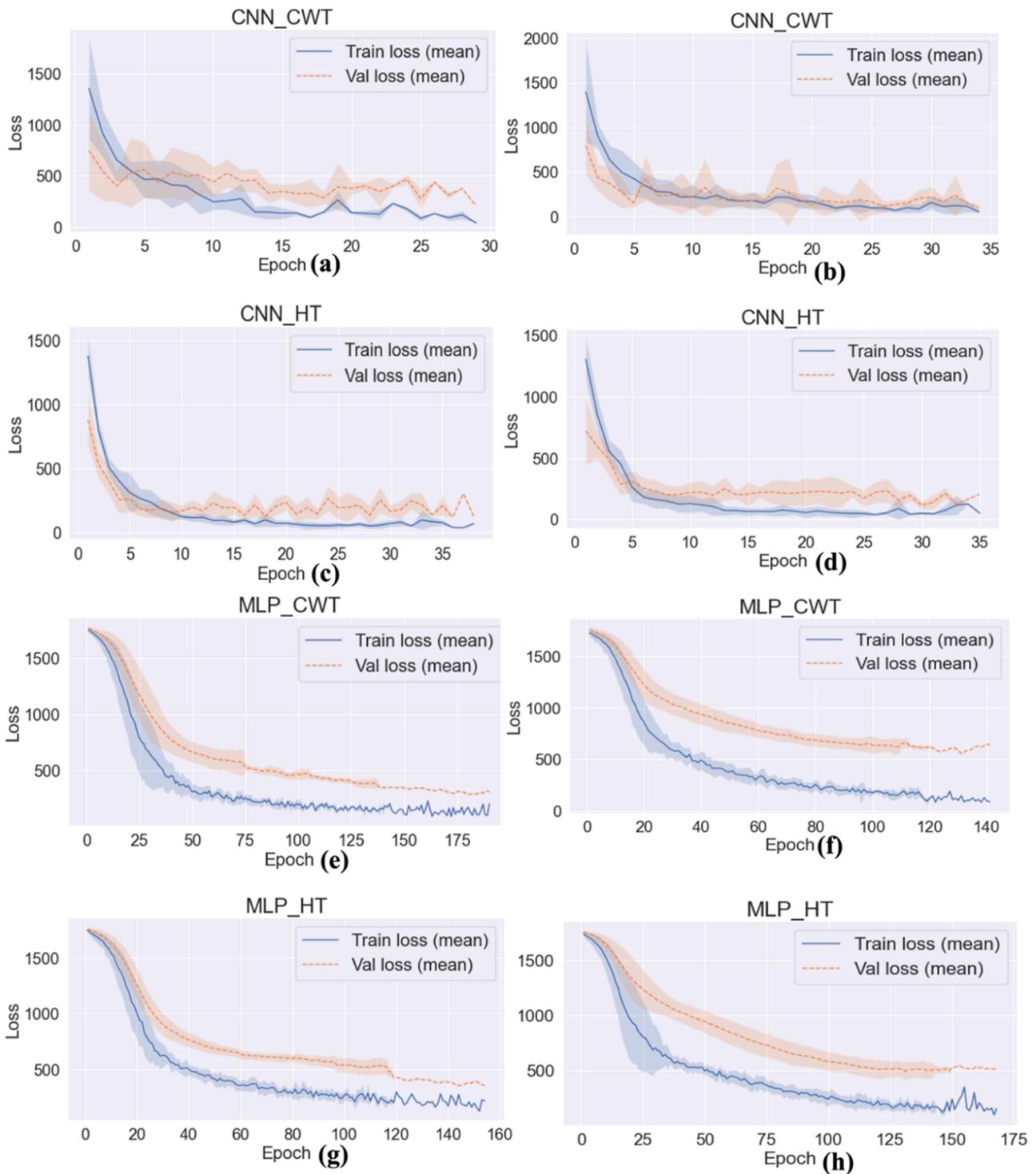
Fig. 16. Learning curves with loss (MSE) for base learners shown over 10 different random initializations of the neural network, considering LOOCV, for fold specimen 2. Subplots (a), (c), (e), and (g) are related to the center excitation frequency (f_{exc}) of 140 kHz, while (b), (d), (f), and (h) are related to the center excitation frequency (f_{exc}) of 160 kHz.

cycles, and other temporal structures, making it a powerful tool for forecasting and analyzing time-dependent data. In time series regression, the 1D-CNN processes the input sequence to extract meaningful features, which are then used to predict future values identifying underlying patterns in the data. The components of full 1D-CNN architecture developed in the present work are given in Fig. 5.

CNN incorporate pooling layers, given in Eqs. (13)-(14), which serve to reduce the spatial dimensions of the data while preserving the most

significant information. Maxpooling is the most common technique, where the maximum value within a defined window is selected. This process not only reduces the computational load but also introduces a form of spatial invariance, making the model less sensitive to small shifts or distortions in the input. Following these layers, the network typically includes fully connected layers, where the high-level features extracted by the convolutional and pooling layers are combined to make final predictions. The entire network is trained end-to-end using

Table 2
MLP architectural parameters as a base learner, an ensemble and a fusion model.

Layer	Type	Activation	Dropout	# Neurons
FCL-1	Fully connected	ReLU	0.2	32
FCL-2	Fully connected	ReLU	—	16
FCL-3	Fully connected	ReLU	—	16
FCL-4	Fully connected	ReLU	—	8
FCL-5	Fully connected	Linear	—	1

Table 3
1D-CNN learner hyperparameters for direct predictor.

Layer	Type	Activation	Dropout	Parameters / # Neurons
Conv1D	Convolution (1D)	ReLU	—	Filters= 128; Kernel Size= 140 (CWT) / 25 (HT)
MaxPooling1D	Pooling	—	—	Pool size= 3
Flatten	Flatten	—	—	—
FCL-1	Fully Connected	ReLU	0.2	128
FCL-2	Fully Connected	ReLU	—	64
FCL-3	Fully Connected	ReLU	—	16
FCL-4	Fully Connected	Linear	—	1

Table 4
L2-LSTM Fusion model architectural parameters.

Layer	Type	Activation	Dropout	# Units / Neurons
LSTM-1	LSTM	ReLU	0.2	32
LSTM-2	LSTM	ReLU	0.2	16
FCL-1	Fully connected	ReLU	0.2	16
FCL-2	Fully connected	ReLU	—	8
FCL-3	Fully connected	Linear	—	1

backpropagation with Adam, allowing it to optimize its filters and weights to minimize the error in predictions.

$$(f*I)(x,y) = \sum_{i=0}^{m} \sum_{j=0}^{n} f(i,j) \cdot I(x+i,y+j) \tag{13}$$

$$p = \max\{x_1, x_2, ..., x_n\} \tag{14}$$

3.3.3. Fusion learners

The final stage incorporates three different fusion strategies: a weighted average ensembling (WAE), a MLP and an LSTM network. The WAE method combines the five RUL-L1 predictions, and its formulation is detailed in Eqs. (15)–(17). WAE is defined as follows:

$$f_{WAE} = \sum_{k=1}^{K} \overline{\omega}_k f_k; \overline{\omega}_k = \frac{\omega_k}{\sum_{k=1}^{K} \omega_k} \tag{15}$$

$$\omega_k^{MSE} = \frac{1}{MSE(Y_{RUL}, RUL^k)} = \frac{1}{\frac{1}{N_j} \sum_{i=1}^{N_j} \left(Y(t_i) - RUL_j^k(t_i)\right)^2} \tag{16}$$

$$\omega_k^{RMSE} = \frac{1}{RMSE(Y_{RUL}, RUL^k)} = \frac{1}{\sqrt{\left[\frac{1}{N_j} \sum_{i=1}^{N_j} \left(Y(t_i) - RUL_j^k(t_i)\right)^2\right]}} \tag{17}$$

where f_k represents the kth individual model and $\overline{\omega}_k$ is its normalized weight. ω_k denotes the weight for the kth individual base model and t_i is the cycle steps for each corresponding RUL point. MSE, RMSE are calculated as error metrics to determine the weights. Simple averaging

ensemble is implemented by setting all ω_k to one. The final output is selected considering the WAE output that has the minimum error with respect to the target RUL value, using only training data.

Fig. 6 shows the LSTM model's learning framework. An LSTM unit utilizes a series of gates to regulate the flow of information into and out of the cell and its corresponding expression is given in the Eqs. (18)–(23):

$$F_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{18}$$

$$I_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{19}$$

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{20}$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \tag{21}$$

$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{22}$$

$$h_t = o_t * \tanh(C_t) \tag{23}$$

Forget Gate (F_t) decides what portion of the cell state should be forgotten. Input Gate (I_t) determines what new information should be stored in the cell state. Output Gate (O_t) decides what the next hidden state should be, which is also used for predictions. It controls how much of the cell state should be output. The cell state is updated based on the input and forget gates. The input gate modulates which new information flows into the cell state, while the forget gate controls the decay of old information. At each time step, the candidate cell state $\widetilde{C}_t$ is computed through a non-linear transformation of the previous hidden state and current input, and the hidden state h_t is then obtained by modulating the updated cell state C_t with the output gate.

3.4. Benchmark machine learning (ML) models

While the proposed multi-level framework quantifies the relative importance and complementary behavior of the feature-frequency combinations within the model, a global assessment of predictive accuracy is presented with external reference models. For this purpose, the classical ML regressors, such as SVR, RF, and AdaBoost, were employed using both DI-Env and DI-CWT features across all excitation frequencies as independent benchmark models.

SVR learns a predictive function by fitting the simplest possible curve that remains within a specified tolerance band around the target values. Only samples that fall outside this band influence the model, becoming support vectors. By using kernel functions such as the radial basis function, SVR can implicitly map inputs into a higher-dimensional space, enabling it to capture nonlinear relationships while maintaining good generalization through a balance between model flatness and error penalization [57].

RF regression builds an ensemble of decision trees; each trained on bootstrapped subsets of the data with randomly selected features. Each tree partitions the feature space into regions and assigns constant predictions to these regions, and the final output is obtained by averaging across all trees. This ensemble structure reduces variance and makes RFs robust against overfitting while effectively modeling nonlinear and interaction-rich feature relationships [58].

AdaBoost regression constructs a weighted ensemble of weak learners trained sequentially. At each iteration, the algorithm increases the emphasis on samples that were previously predicted poorly, forcing subsequent learners to focus more on difficult cases. Each weak learner contributes to the final model with a weight proportional to its accuracy, and the overall prediction becomes a weighted sum of all learners. This iterative reweighting mechanism reduces bias and improves predictive power, although it can make the model more sensitive to noisy data [59].
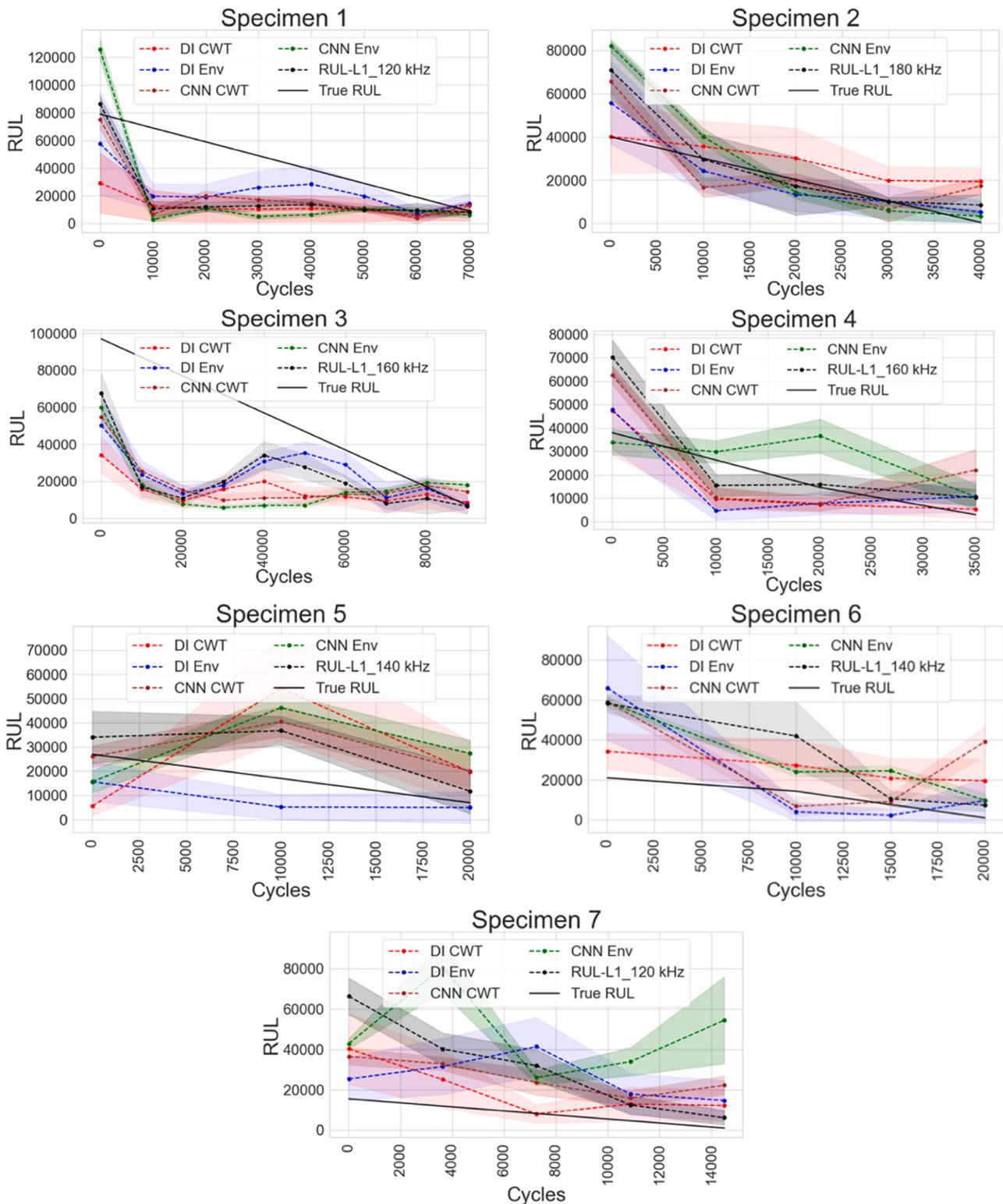
Fig. 17. Base learner prediction results for the excitation frequency of highest accuracy with their Level-1 results.

4. Compression after impact (CAI) fatigue testing: experimental setup

The experimental campaign was conducted on woven-type CFRP plates with a thickness of 5.5 mm and a layup of $[\pm45/0/90/\mp45]s$, as shown in Fig. 7. The specimens, manufactured in accordance with ASTM D7136 [60] for CAI testing, had dimensions of 100 mm in width and 150 mm in length. Controlled low-velocity impacts were introduced using a drop-weight tower to evaluate their effect on fatigue performance. The experimental setup is illustrated in Fig. 8, comprising the fatigue testing configuration, an anti-buckling fixture to control global buckling, and the GW unit for in-situ data collection, and the impact testing setup with its principal components. The complete dataset, including details of the C-C fatigue after impact tests, has been published and is available in [39]. The GW unit incorporated instruments tailored for SHM applications: a Lecroy HD0634 oscilloscope for signal acquisition, an Agilent

Table 5
Base learner MAPE values for different f_{exc} and processed inputs. The values are obtained after averaging over the time step.

Excitation frequencies		Input types			
		DI-CWT	DI-Env	CNN-CWT	CNN-Env
100 kHz	Specimen 1	0.713	0.688	**0.640**	0.787
	Specimen 2	12.666	15.877	**5.453**	12.992
	Specimen 3	**0.637**	0.643	0.640	0.736
	Specimen 4	1.712	1.395	**0.482**	2.302
	Specimen 5	0.399	**0.169**	1.633	1.082
	Specimen 6	**3.882**	9.601	4.333	7.727
	Specimen 7	3.877	**1.859**	4.261	13.486
	Average	3.412	4.319	**2.492**	5.587
120 kHz	Specimen 1	0.646	**0.488**	0.591	0.695
	Specimen 2	7.478	6.853	5.704	**4.250**
	Specimen 3	0.653	**0.549**	0.652	0.897
	Specimen 4	0.644	1.013	**0.563**	2.085
	Specimen 5	0.331	**0.165**	1.316	2.348
	Specimen 6	3.768	**3.082**	3.345	5.236
	Specimen 7	**2.407**	3.174	4.890	3.719
	Average	2.275	**2.189**	2.437	2.747
140 kHz	Specimen 1	0.708	**0.485**	0.596	0.603
	Specimen 2	**3.565**	8.086	11.091	11.084
	Specimen 3	**0.598**	0.740	0.683	0.706
	Specimen 4	2.120	**0.354**	1.205	2.463
	Specimen 5	1.690	**0.488**	1.153	1.663
	Specimen 6	5.444	**3.055**	10.171	3.388
	Specimen 7	2.213	7.083	4.955	**3.144**
	Average	**2.334**	2.899	4.265	3.293
160 kHz	Specimen 1	0.639	0.718	**0.591**	0.714
	Specimen 2	3.859	**2.570**	4.373	6.562
	Specimen 3	0.630	**0.435**	0.662	0.751
	Specimen 4	**0.319**	0.468	0.856	0.417
	Specimen 5	0.649	0.741	**0.491**	1.547
	Specimen 6	4.517	**1.054**	4.976	5.688
	Specimen 7	2.429	4.575	**2.298**	11.623
	Average	1.863	**1.509**	2.035	3.900
180 kHz	Specimen 1	0.657	0.738	0.690	**0.627**
	Specimen 2	7.870	2.095	7.066	**1.543**
	Specimen 3	0.653	**0.583**	0.636	0.744
	Specimen 4	**0.607**	1.147	2.346	1.299
	Specimen 5	1.047	0.462	0.595	**0.351**
	Specimen 6	5.444	**3.055**	10.171	3.388
	Specimen 7	**3.146**	4.574	5.724	13.906
	Average	2.775	**1.808**	3.890	3.123

33500B signal generator for excitation, and a KEYSIGHT 34972A multiplexer for channel switching. Impedance measurements were performed using a KEYSIGHT E4980A analyzer, and all components were integrated within a National Instruments PXIe-1071 to ensure synchronized control and data management. Fig. 9 illustrates the sensor network used for the GW application, consisting of six PZT transducers linearly distributed in an array along the top and bottom sections of the specimen and operated in a pitch–catch configuration. LOCTITE EA 9394 AERO [61] is a two-part structural paste adhesive used to attach PZTs on the surface of the CFRP test samples. PIC255 type PZT disk transducers have been used in this experiment with a diameter of 8 mm, thickness 0.5 mm, with diameter and thickness frequency constants Np and Nt; 2000 and 1420, respectively, where further information can be found in [62]. The length of the soldered cables is adjusted to 50 cm. GW excitation is applied at center frequencies f_{exc} of 100, 120, 140, 160, and 180 kHz using 10V 2-cycle tone-burst signal.

Fatigue testing starts with a slow cycle with a 1 Hz frequency, and later cycles are executed with a 5 Hz frequency with a load ratio of 10. Fatigue testing was initiated at a loading frequency of 1 Hz to ensure stable load introduction and reliable monitoring during the initial loading cycles. Once stable cyclic behavior was confirmed, the loading frequency was increased to 5 Hz to reduce the overall test duration. This frequency change was not triggered by a specific damage indicator or delamination threshold and did not modify the load ratio or influence the governing damage and failure mechanisms. In Fig. 10, the fatigue

force plot is shown. Constant and non-constant fatigue conditions were explored within the dataset, and their details are given in Table 1. During the fatigue testing, the test is paused at lower force levels to enable the operation of ultrasonic C-scan monitoring, which tracks the progression of delamination in parallel with GW acquisition.

In Fig. 11, C-scan-based delamination length quantification is presented, serving as the degradation indicator while specimens #4 and #5 are excluded due to the noise present in their C-scan images, which prevented the accurate quantification of delamination length. Fig. 12 illustrates the experimentally measured delamination evolution under C-C fatigue loading in which (a) shows the maximum delamination length extracted from C-scan inspections, while (b) presents the corresponding normalized growth rates. In Fig. 12, the progression of delamination is presented in terms of measured length, and the growth rate is calculated considering Eq. (24):

$$\text{Normalized Growth Rate} = \frac{L_N - L_1}{L_1} \qquad (24)$$

L_N is delamination length measured at a given cycle N and L_1 is initial delamination length measured at the first cycle. In contrast to the simplified schematic representation given in Fig. 1, which was introduced to conceptually define the RUL estimation framework, the experimental observations reveal that delamination growth is neither continuous nor purely exponential. Instead, the behavior is highly stochastic: the delamination length often remains nearly stable during the initial fatigue cycles and then exhibits a pronounced acceleration at later cycles. This acceleration effect is consistent with the expected mechanics of compression-after-impact, where local buckling and interaction mechanisms drive a rapid propagation of delamination once a critical condition is reached.

5. Prognostic results

5.1. GW-DIs and delamination severity

GW-DIs are derived using signal processing techniques to isolate segments of the GW signal that are particularly aimed to be sensitive to delamination (see Fig. 13).

In Fig. 14, the temporal evolution of the DI-Env and DI-CWT across the fatigue life are shown for seven specimens at each f_{exc}, with values averaged over a total of 18 sensor-actuator paths. Each group of DIs was represented by a single color, covering all corresponding f_{exc}. Here, time steps refer to individual data acquisition measurement steps, which are specific to each specimen. DIs generally exhibit an increasing trend with time showing consistent progression across specimens.

Fig. 15 quantifies the correlation between the DIs and the maximum delamination length with Corr function across different f_{exc} values, where f1 represents 100 kHz, followed by 120 kHz, 140 kHz, 160 kHz, and 180 kHz up to f5. The results indicate that DI-CWT generally achieves higher correlation values across most frequency components and specimens, particularly in Specimens #6 and #7. DI-Env, while contributing to overall correlation, shows more pronounced variability across frequencies. Both DI-Env and DI-CWT exhibit with lower correlation in Specimens #1 and #3.

5.2. Prognostic model training & testing

Training a DNN model involves adjusting neuron weights and biases to minimize a loss function, which measures the difference between predicted outputs and actual labels which is achieved using Adam in this study [63]. Dropout is a regularization technique [64] used to mitigate overfitting in artificial neural networks by reducing the risk of complex co-adaptations among neurons during training. To ensure tailored feature extraction, CNN kernels are explicitly tuned for each input. At each f_{exc}, the processed GW signals are passed through the base learners,
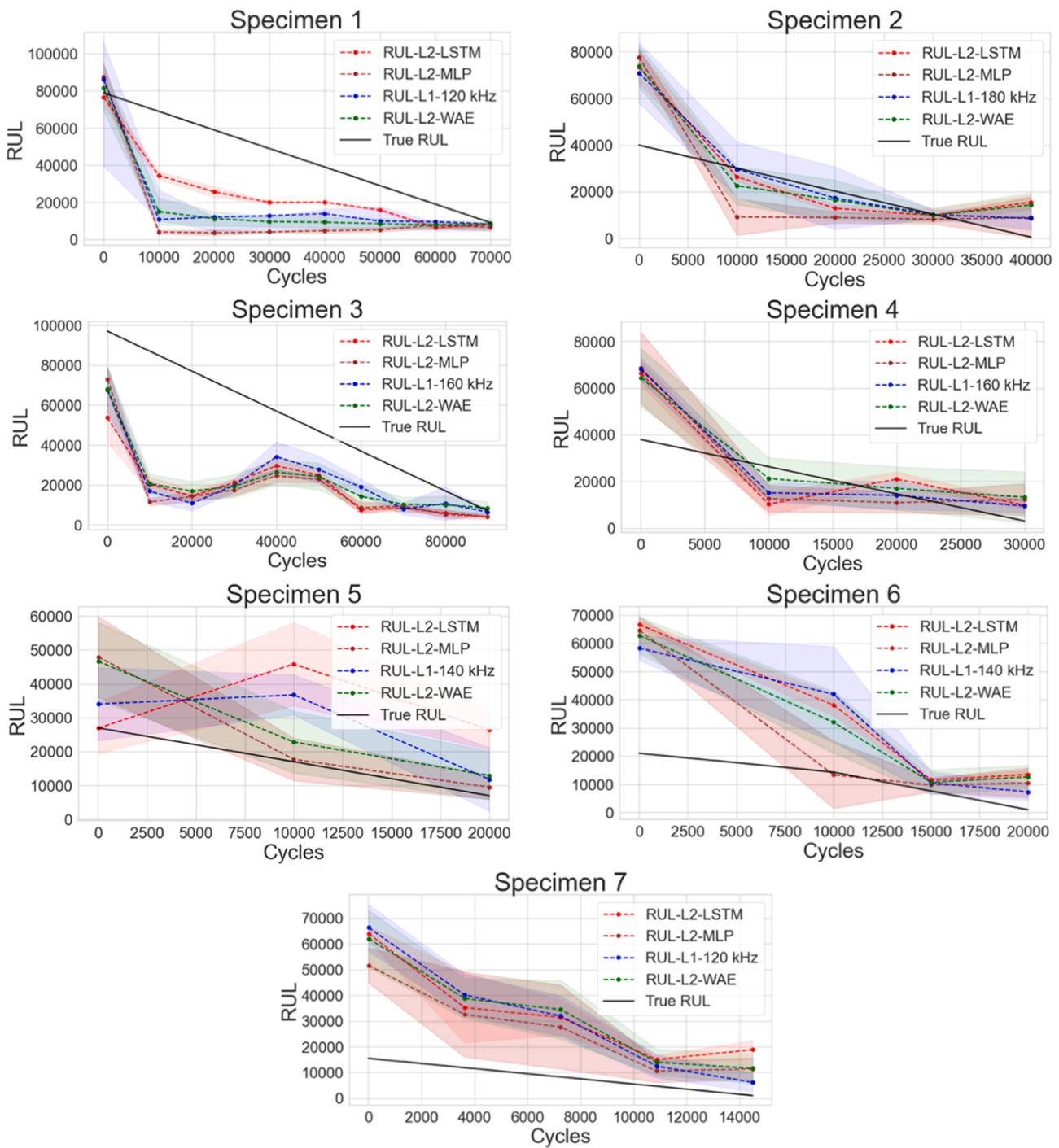
Fig. 18. RUL prediction from different levels (L1 and L2) composed of different models.

combined at the ensemble level, and fused across frequencies. The model is trained using input data structured as a time series, but without incorporating explicit time dependence within its base-learner architecture. The input datasets are formatted as $N_{cycle\ steps} \times N_{features}$ for MLP model and $N_{cycle\ steps} \times L_{signal\ length} \times N_{features}$ for 1D-CNN model both split into testing and training sets. The former dataset has 39×18 and the dataset used to train for 1D-CNN has $39 \times 1401 \times 18$ for CNN CWT and $39 \times 250 \times 18$ for CNN Env for the first fold.

A leave-one-out cross-validation (LOOCV) strategy is adopted in this study. In each fold, six specimens are used for training, while the remaining specimen is completely excluded from training and reserved for testing. For example, in fold 1, specimen 1 is held out as the test

specimen, and the other six specimens constitute the training dataset. The test specimen choice rotates through all available specimens for different folds. During testing within each fold, the model predicts the RUL at each time step using degradation data from the held-out specimen, which is unseen during training in that fold, and iteratively estimates RUL over its entire time horizon. Thus, although each specimen is eventually used for both training and testing across different folds, no specimen is ever used for training and testing simultaneously within the same fold.

Input data are standardized using z-score normalization, applied separately to each DI set, sensor-actuator path under each f_{exc}. Hyperparameters were selected through an empirical tuning process guided by validation performance measured as the mean squared error (MSE). The

Table 6
Averaged MAPE error values for RUL prediction from each fusion level.

		Spec. 1	Spec. 2	Spec. 3	Spec. 4	Spec. 5	Spec. 6	Spec. 7	Average
Fusion levels	L1-100kHz	0.610	14.611	0.582	1.139	**0.469**	6.727	5.571	4.244
	L1-120kHz	**0.543**	7.085	0.582	1.351	1.186	3.587	**3.073**	2.487
	L1-140kHz	0.596	10.466	0.549	1.379	0.601	**2.599**	4.034	2.889
	L1-160kHz	0.573	3.884	0.527	**0.764**	0.864	3.868	4.496	2.139
	L1-180kHz	0.620	**3.396**	**0.525**	1.639	0.582	5.198	5.631	2.513
	L2-WAE	0.568	5.785	0.543	1.113	0.639	3.825	4.214	2.384
	L2-MLP	0.685	3.855	0.626	1.169	0.392	2.984	3.614	*1.904*
	L2-LSTM	0.426	6.255	0.621	1.014	1.493	4.212	5.608	2.804
Benchmark models	**AdaBoost**	0.477	4.953	0.591	1.198	1.324	4.494	4.328	2.480
	Random F.	0.460	5.432	0.548	2.009	1.218	5.658	5.464	2.969
	SVR	0.551	6.758	0.622	1.359	0.870	6.544	5.563	3.181

number of neurons per layer and other capacity-related parameters were varied systematically, and the validation loss was monitored across configurations. It was observed that further increases in model capacity produced no meaningful reduction in validation error beyond a certain threshold, indicating convergence of the architecture's representational ability. The final configurations were therefore fixed at the smallest architecture that ensured stable performance while avoiding over-parameterization. To address the stochastic variability inherent in neural network training, weights were randomly initialized for each run, and the models were trained ten times per LOOCV fold. Fold-level results are reported as the mean and standard deviation in Fig. 16 across these repeated runs, ensuring both robustness and reproducibility of the reported performance. Validation data is randomly selected from the training set, with 20% of the data set aside for validation purposes and to monitor the training loss. Early stopping is employed during training to limit excessive fitting and improve generalization. Although it does not directly modify the loss function or model parameters, early stopping terminates training once the validation loss no longer improves, preventing the model from adapting to noise or non-generalizable patterns. Table 2 presents the parameters for each fully connected layer (FCL) in the MLP learners and Table 3 presents the parameters of 1D-CNN base learner.

Although all base learners shown in Fig. 16 are trained on datasets of identical size, the amount and richness of degradation information they receive differ across input representations and excitation frequencies. Signal-based models operate on full GW waveforms, which provide dense physical information such as dispersion, mode coupling, and phase/amplitude variations. DI-based learners, in contrast, rely on low-dimensional engineered features that summarize only specific aspects of the signal, potentially missing subtle frequency-dependent degradation trends. As a result, differences in generalization behavior emerge. CNNs can learn from the raw temporal complexity of the signal, while DI-based MLPs may be more susceptible to overfitting due to limited input expressiveness. While validation loss during training did not consistently reflect this, the small dataset size and randomly split validation sets may have produced optimistic generalization estimates, particularly for low-capacity models with compressed inputs. The proposed ensemble and fusion architecture naturally mitigate these limitations by integrating complementary predictions across models and frequencies, reinforcing stable, damage-relevant patterns in the final RUL estimation.

In parallel, the MLP and LSTM models are directly connected to the initial 20 outputs of the base learners. Both models receive all 20 base predictions as input, with the LSTM model implemented using a 3D input format (sample size × time step × features), where the time step is set to 1 and the sample size dynamically adjusted during each LOOCV. LSTM model hyperparameters are given in Table 4.

All benchmark models (i.e., SVR, RF, and AdaBoost) were trained using the same feature set, obtained by concatenating CWT-based and HT-based GW-DIs into a fused DI representation. The dataset was split into training and test sets corresponding to the LOOCV methodology: 6 specimens for training and 1 specimen for testing iteratively. Each model was trained independently for 10 repeated runs, each initialized with a different random seed to assess variance and model stability. Models were trained using fixed configurations, where 300 estimators for RF, 100 for AdaBoost, and $C = 10$, $\epsilon = 0.1$ for SVR were considered.

5.3. Testing results & discussion

Fig. 17 presents the predicted RUL for all specimens based on LOOCV folds using different models and input types, including DI-CWT, DI-Env, CNN-CWT, and CNN-HT, results in RUL-1, RUL-2, RUL-3 and RUL-4 for all f_{exc}. The curves represent the predicted RUL trajectories compared with the true RUL, with shaded areas indicating the confidence bounds. Table 5 presents average error values for each base learner model and input types of each sample. Eq. (25) calculates the error value using Mean Absolute Percentage Error (MAPE) function where Y_i^p for the predicted value of RUL and Y_i^m is the true label for RUL:

$$MAPE = \frac{1}{N_{cycles}} \sum_{i=1}^{N_{cycles}} \left| \frac{Y_i^m \; Y_i^p}{Y_i^m} \right| \qquad (25)$$

The correlations presented in Fig. 15 provide valuable insights into the relationship between the measured delamination length and the DIs. While a one-dimensional quantified delamination length serves as a useful reference for assessing damage progression, GW signals may contain richer information about the damage area. Consequently, the correlation of GW-DIs with delamination length may appear strong, yet this does not necessarily translate into a proportionally strong contribution to RUL estimation. The reason for this may be that delamination-related information is well isolated in the signal, whereas other degradation mechanisms, such as matrix cracks or additional damage types, are embedded in the full signal through effects like mode conversion. As a result, DI-based calculations may remain limited in capturing the broader spectrum of damage evolution and thus in predicting RUL. For instance, both DIs exhibit high correlation values for Specimen #7; however, as shown in Fig. 17, the base CNN models may demonstrate higher estimation accuracy compared to models using DI-based inputs. This suggests that the CNN model is better equipped to capture degradation indicators beyond delamination length, potentially accounting for varied damage types such as matrix cracks, and thereby providing a more accurate RUL prediction.

For Specimens #1 and #3, the correlation is observed to be the lowest, as shown in Fig. 15 and the base learner results indicate the lowest prediction accuracy during the earlier cycles yet converges in later cycles. According to the Fig. 12, Specimens #1 and #3 exhibit slower degradation in their earlier cycles compared to the other specimens in the dataset. This observation may suggest that DIs may be less effective in capturing the early stages of delamination progression for these specimens, while improved accuracy is observed in the base learner model during the later cycles for Specimen #1 and #3. DI-CWT presents higher score in terms of its correlation with measured delamination length; the base learner prediction results indicate that DI-Env

has higher accuracy in RUL predictions despite its lower correlation score with C-scan based measured delamination length. This highlights that the DI-Env feature indicated greater severity than the C-scan measurements, provided a more representative description of damage progression, and thereby contributed more effectively to RUL prediction.

In Fig. 17, Level-1 demonstrates robustness in combining the strengths, minimizing large deviations seen in individual models, as observed in Specimen #4, #5, and #6. Especially, in the result of Specimen #7, Level-1 presents higher performance reducing the error of each model giving a better convergence in the final cycle. As shown in Table 5, the CNN-CWT model achieves the highest average accuracy at the lower f_{exc} of 100 kHz. In contrast, at higher frequencies, models based on GW-DIs exhibit lower error values, with the DI-Env model performing particularly well at 160 kHz and 180 kHz. This may suggest that the captured GW mode in DI calculation at these higher frequencies exhibits a stronger interaction with existing structural damage that contributes more significantly to failure. While average error rates may suggest that a particular model performs optimally at a specific f_{exc}, the performance of each model varies across different specimens at each frequency level. The variability in model performance across different specimens indicates that no single method consistently excels, emphasizing the need for a complementary phase capable of generalizing across diverse degradation characteristics.

Fig. 18 shows the RUL prediction results for all specimens for ensemble and fusion stages. The L1 ensemble model output is presented for the best-performing f_{exc} (160 kHz), while the L2 fusion results are provided for both MLP- and LSTM-based prognostic models. Table 6 indicates the average MAPE error for RUL predictions produced from each level. The error metric was calculated and averaged across cycles to evaluate overall performance. Level-1 show some strong results, their higher errors in various specimens and inputs indicate that they struggle to generalize across all conditions. Despite RUL-L1-160 kHz performing well across samples with an average error of 2.139, the L2-MLP model outperforms all with the lowest average error of 1.904. The L2-WAE model also performs strongly, with an average error of 2.384.

The LSTM and MLP models show higher adaptability by leveraging more complex, multi-dimensional input information. L2-MLP in particular stands out, delivering the best average performance across all samples. Its ability to capture complex degradation trends suggests that MLP effectively generalizes better than LSTM, which shows higher error rates. The L2-MLP model emerges as the most effective approach, achieving the lowest average error across all samples. Its ability to integrate multi-frequency information and adapt to varying signal dynamics enables it to handle the inherent complexity in GW signals more effectively than single-frequency models. Similarly, the L2-WAE model demonstrates robust performance, achieving low errors across multiple samples due to its ability to balance contributions from all excitation frequencies, thus mitigating the limitations of single-frequency models. This model excels at stabilizing predictions by smoothing out fluctuations, providing reliable RUL estimates across a range of conditions. Despite these strengths, proposed models exhibit some limitations that need to be addressed. L2-LSTM, while highly adaptable in capturing temporal dependencies, shows greater prediction uncertainty in certain samples, as evidenced by the wider confidence intervals observed in cases such as Specimen #5. This higher uncertainty may stem from the model's sensitivity to minor fluctuations in the data, which can result in overfitting, particularly when the degradation trajectory includes abrupt changes or a limited number of time steps. LSTM's focus on temporal dependencies can lead to increased prediction error, while MLP, with its simpler structure, generalizes better in this scenario with 32% higher performance than LSTM model. Furthermore, while L2-WAE provides a more stable and balanced prediction by aggregating information from multiple excitation frequencies, it may struggle in scenarios where high error arises from averaged predictions, as it is an explicit model that may oversimplify complex degradation patterns by smoothing out important variations.

6. Conclusion

This study proposed a multi-level, frequency-aware prognostic framework for predicting the RUL of impacted CFRP structures under fatigue loading. By integrating direct data-driven models with feature-based approaches the framework effectively captured degradation patterns across multiple excitation frequencies. This hybrid strategy improved both predictive accuracy and enabled interpretable RUL results, demonstrating that the combination of direct learning and feature engineering provides complementary strengths for fatigue prognosis. The results showed that while single-frequency models can perform well in specific conditions, they fail to generalize across complex delamination scenarios. Multi-frequency fusion, particularly with the L2-MLP model achieving the lowest overall error, average MAPE 1.904, corresponding to an improvement of 11–55% compared with the best and worst single-frequency baseline predictions and 30% higher accuracy compared to the lowest error benchmark ML model. These findings highlight the critical role of integrating information from multiple excitation frequencies to achieve robust and reliable RUL prognosis in impacted CFRP laminates under compressive fatigue loading.

While the proposed multi-level DL framework demonstrated promising results for RUL prediction for delamination induced failure, several limitations for the whole framework should be acknowledged. First, the experimental dataset was limited in size, comprising seven specimens with varying damage severities and loading histories. Although the number of tested samples is adequate for demonstrating the proposed framework, broader validation across structural scenarios involving multiple impacts or varying damage locations would be necessary to confirm its generalizability in more complex real-world conditions. Secondly, delamination severity was approximated using a one-dimensional projection of delamination length obtained from C-scan measurements. Although this simplified metric provides a consistent and quantifiable reference that supports the interpretation of the derived DIs and enables comparative analysis across samples, it does not fully reflect the complex, three-dimensional nature of delamination. To improve physical interpretability, more physics-based considerations of GW–delamination interactions should be incorporated into the signal processing stage, enabling DIs to more directly capture the mechanisms of damage accumulation. Ultimately, this also highlights the broader need for refined physical–mechanical analyses of delamination evolution in CFRPs to complement data-driven approaches and enhance the reliability of RUL prognosis.

CRediT authorship contribution statement

Ferda C. Gül: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Morteza Moradi:** Writing – review & editing, Writing – original draft, Validation, Methodology, Conceptualization, Visualization. **Dimitrios Zarouchas:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Data availability

The dataset used in this article is available at https://doi.org/10.4121/BB64D43C-2ECC-4841-9744-A2377CAC9167.

References

[1] K.I. Tserpes, P. Papanikos, G. Labeas, S. Pantelakis, Fatigue damage accumulation and residual strength assessment of CFRP laminates, Compos. Struct. 63 (2) (2004) 219–230, https://doi.org/10.1016/S0263-8223(03)00169-7.

[2] A. Wan, J. Xiong, Y. Xu, Fatigue life prediction of woven composite laminates with initial delamination, Fatigue Fract. Eng. Mater. Struct. 43 (9) (2020) 2130–2146, https://doi.org/10.1111/ffe.13296. Sep.

[3] D. Biagini, J.A. Pascoe, R. Alderliesten, Investigating apparent plateau phases in fatigue after impact damage growth in CFRP with ultrasound scan and acoustic emissions, Int. J. Fatigue 177 (2023), https://doi.org/10.1016/j.ijfatigue.2023.107957. Dec.

[4] F. Lamon, L. Maragoni, P.A. Carraro, M. Quaresimin, Fatigue damage evolution in woven composites with different architectures, Int. J. Fatigue 167 (2023) 107365, https://doi.org/10.1016/j.ijfatigue.2022.107365. Part BFeb.

[5] European Union Aviation Safety Agency, Annex to ED Decision 2022/001/R, General Acceptable Means of Compliance for Airworthiness of Products, Parts and Appliances (AMC-20), Amendment 23 (2022), 21 January 2022.

[6] C. Rans, R.C. Alderliesten, R. Benedictus, Misinterpreting the results: how similitude can improve our understanding of fatigue delamination growth, Compos. Sci. Technol. 71 (2) (2011) 230–238, https://doi.org/10.1016/j.compscitech.2010.11.010. Jan.

[7] R.C. Alderliesten, Designing for damage tolerance in aerospace: a hybrid material technology, Mater. Des. 66 (2015) 421–428, https://doi.org/10.1016/j.matdes.2014.07.059. Feb.

[8] M. Corbetta, C. Sbarufatti, M. Giglio, A. Saxena, K. Goebel, A Bayesian framework for fatigue life prediction of composite laminates under co-existing matrix cracks and delamination, Compos. Struct. 187 (2018) 58–70, https://doi.org/10.1016/j.compstruct.2017.12.035. Mar.

[9] J.A. Pascoe, R.C. Alderliesten, R. Benedictus, Methods for the prediction of fatigue delamination growth in composites and adhesive bonds – a critical review, Eng. Fract. Mech. 112–113 (2013) 72–96, https://doi.org/10.1016/j.engfracmech.2013.10.003. Nov.

[10] D.J. Elder, R.S. Thomson, M.Q. Nguyen, M.L. Scott, Review of delamination predictive methods for low-speed impact of composite laminates, Compos. Struct. 66 (1–4) (2004) 677–683, https://doi.org/10.1016/j.compstruct.2004.06.004. Oct.

[11] C. Ferreira, G. Gonçalves, Remaining useful life prediction and challenges: a literature review on the use of machine learning methods, J. Manuf. Syst. 63 (2022) 550–562, https://doi.org/10.1016/j.jmsy.2022.04.004. Apr.

[12] N. Eleftheroglou, T. Loutas, Fatigue damage diagnostics and prognostics of composites utilizing structural health monitoring data and stochastic processes, Struct. Health Monit. 15 (4) (2016) 473–488, https://doi.org/10.1177/1475921716646579. Jul.

[13] G. Galanopoulos, N. Eleftheroglou, D. Milanoski, A. Broer, D. Zarouchas, T. Loutas, A novel strain-based health indicator for the remaining useful life estimation of degrading composite structures, Compos. Struct. 306 (2023), https://doi.org/10.1016/j.compstruct.2022.116579. Feb.

[14] M. Philibert, K. Yao, M. Gresil, C. Soutis, Lamb waves-based technologies for structural health monitoring of composite structures for aircraft applications, Eur. J. Mater. 2 (1) (2022) 436–474, https://doi.org/10.1080/26889277.2022.2094839. Dec.

[15] Q. Hu, C. Xu, J. Zhao, Y. Zhang, R. Liu, Z. Zhang, N. Li, W. Wang, A novel framework for ultrasonic guided wave-based structural health monitoring of CFRP-reinforced steel structures under varying temperatures, Thin-Walled Struct. 216 (2025) 113716, https://doi.org/10.1016/j.tws.2025.113716. Part BNov.

[16] A. De Luca, F. Caputo, Z.Sharif Khodaei, M.H. Aliabadi, Damage characterization of composite plates under low velocity impact using ultrasonic guided waves, Compos. B Eng. 138 (2018) 168–180, https://doi.org/10.1016/j.compositesb.2017.11.042. Apr.

[17] Z. Su, L. Ye, Y. Lu, Guided Lamb waves for identification of damage in composite structures: a review, J. Sound. Vib. 295 (3–5) (2006) 753–780, https://doi.org/10.1016/j.jsv.2006.01.020. Aug.

[18] M. El Mountassir, S. Yaacoubi, S. Dellagi, M. Sfar, M. Aouini, An ultrasonic guided waves based prognostic approach for predictive maintenance: Experimental study cases, in: Mechanical Systems and Signal Processing 190, 2023, https://doi.org/10.1016/j.ymssp.2023.110135.

[19] N.Testoni Marzani, L. De Marchi, M. Messina, E. Monaco, A. Apicella, An open database for benchmarking guided waves structural health monitoring algorithms on a composite full-scale outer wing demonstrator, Struct. Health Monit. 18 (5–6) (2019) 1–18, https://doi.org/10.1177/1475921719889029.

[20] H. Mei, V. Giurgiutiu, Characterization of multilayer delamination in composites using wavenumber analysis: numerical and experimental studies, Struct. Health Monit. 19 (6) (2020) 1–26, https://doi.org/10.1177/1475921720939616.

[21] R.P. Palanisamy, P. Banerjee, S. Mukherjee, M. Haq, Y. Deng, Fatigue damage prognosis in adhesive bonded composite lap-joints using guided waves, in: Proceedings of the Annual Conference of the Prognostics and Health Management

[22] F. Hervin, P. Fromme, Anisotropy influence on guided wave scattering for composite structure monitoring, Struct. Health Monit. 22 (4) (2023) 2626–2640, https://doi.org/10.1177/14759217221133284.

[23] J. Chen, S. Yuan, L. Qiu, H. Wang, W. Yang, On-line prognosis of fatigue crack propagation based on Gaussian weight-mixture proposal particle filter, Ultrasonics. 82 (2018) 134–144, https://doi.org/10.1016/j.ultras.2017.07.016.

[24] Z. Li, J. Jia, M. Wang, M. Gu, S. Tu, Particle filter for fatigue crack growth prediction using SH0 wave on-line monitoring, Ultrasonics 142 (2024) 107355, https://doi.org/10.1016/j.ultras.2024.107355.

[25] S. Mishra, O.A. Vanli, Remaining useful life estimation with lamb-wave sensors based on wiener process and principal components regression, Struct. Health Monit. 15 (5) (2016) 473–488, https://doi.org/10.1177/1475921715624503.

[26] M.R. Abir, T.E. Tay, M. Ridha, H.P. Lee, On the relationship between failure mechanism and compression after impact (CAI) strength in composites, Compos. Struct. 182 (2017) 242–250, https://doi.org/10.1016/j.compstruct.2017.09.038. Dec.

[27] Z. Yang et al., A review on guided-ultrasonic-wave-based structural health monitoring: From fundamental theory to machine learning techniques, Ultrasonics, 133, 107014, Aug. 2023, 10.1016/j.ultras.2023.107014.

[28] L. Lomazzi, M. Giglio, F. Cadini, Towards a deep learning-based unified approach for structural damage detection, localisation and quantification, Eng. Appl. Artif. Intell. 121 (2023), https://doi.org/10.1016/j.engappai.2023.106003. May.

[29] R. Geng, H. Zhang, H. Li, Machine learning-based structural health monitoring and prognostics: Emerging trends and applications, Compos. Sci. Technol. 292 (2025) 110771, https://doi.org/10.1016/j.compscitech.2025.110771. Jan.

[30] E.Ozer Malekloo, M. AlHamaydeh, M. Girolami, Machine learning and structural health monitoring overview with emerging technology and high dimensional data source highlights, Struct. Health Monit. 21 (4) (2022) 1906–1955, https://doi.org/10.1177/14759217211036880.

[31] S. Ullah, A.A. Ijjeh, P. Kudela, Deep learning approach for delamination identification using animation of Lamb waves, Eng. Appl. Artif. Intell. 117 (2023), https://doi.org/10.1016/j.engappai.2022.105520. Jan.

[32] G. Galanopoulos, E. Fytsilis, N. Yue, A. Broer, A data-driven methodology for upscaling remaining useful life predictions: From single to multistiffened composite panels, Compos. Part C Open Access 11 (2023) 100366, https://doi.org/10.1016/j.jcomc.2023.100366.

[33] J. Xia, Y. Feng, C. Lu, C. Fei, X. Xue, LSTM-based multi-layer self-attention method for remaining useful life estimation of mechanical systems, Eng. Fail. Anal. 125 (Jul. 2021), https://doi.org/10.1016/j.engfailanal.2021.105385.

[34] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, D. Inman, 1D convolutional neural networks and applications: a survey, Mech. Syst. Signal Process. 151 (2021), https://doi.org/10.1016/j.ymssp.2020.107398. April.

[35] B. Feng, S. Cheng, K. Deng, Y. Kang, Localization of low-velocity impact in CFRP plate using time–frequency features of guided wave and convolutional neural network, Wave Motion. 119 (2023), https://doi.org/10.1016/j.wavemoti.2023.103127. Jun.

[36] M. Moradi, F.C. Gul, D. Zarouchas, A novel machine learning model to design historical-independent health indicators for composite structures, Compos. B Eng. 275 (2024), https://doi.org/10.1016/j.compositesb.2024.111328. Apr.

[37] F.C. Gül, M. Moradi, D. Zarouchas, Active sensing-based prognostics for impacted CFRP structures under compressive fatigue loading, Struct. Health Monit. (2025) 1–22, https://doi.org/10.1177/14759217251317747.

[38] M. He, Y. Wang, K.R. Ramakrishnan, Z. Zhang, A comparison of machine learning algorithms for assessment of delamination in fiber-reinforced polymer composite beams, Struct. Health Monit. (2020) 1–16, https://doi.org/10.1177/1475921720967157.

[39] F. Gül, D. Zarouchas, C-C Fatigue Testing for Impacted Woven-type CFRP Samples, 4TU.ResearchData, 2024, https://doi.org/10.4121/BB64D43C-2ECC-4841-9744-A2377CAC9167.V1.

[40] A. Sattarifar, T. Nestorović, Emergence of machine learning techniques in ultrasonic guided wave-based structural health monitoring: a narrative review, Tech. Rep. (2023) [Online].

[41] Z. Yang, H. Yang, T. Tian, D. Deng, M. Hu, J. Ma, D. Gao, J. Zhang, S. Ma, L. Yang, H. Xu, Z. Wu, A review on guided-ultrasonic-wave-based structural health monitoring: From fundamental theory to machine learning techniques, Ultrasonics. 133 (2023) 107014, https://doi.org/10.1016/j.ultras.2022.107014. JanArt. no.

[42] J.K. Kim, M.L. Sham, Impact and delamination failure of woven-fabric composites, Compos. Sci. Technol. 62 (5–6) (2002) 655–661, https://doi.org/10.1016/S0266-3538(02)00005-8.

[43] V.S. Anuse, S.K.R. Vellmurugan, S.K. Ha, Compression-after-impact analysis of carbon/epoxy and glass/epoxy hybrid composite laminate with different ply orientation sequences, Thin-Walled Struct. 185 (2023) 110608, https://doi.org/10.1016/j.tws.2023.110608. Apr.

[44] D. Hull, Y.B. Shi, Damage mechanism characterization in composite damage tolerance investigations, Compos. Struct. 23 (2) (1993) 99–120, https://doi.org/10.1016/0263-8223(93)90113-R.

[45] T. Ogasawara, S. Sugimoto, H. Katoh, T. Ishikawa, Fatigue behavior and lifetime distribution of impact-damaged carbon fiber/toughened epoxy composites under compressive loading, Adv. Compos. Mater. 22 (2) (2013) 65–78, https://doi.org/10.1080/09243046.2013.768324.

[46] A. Sobhani, M. Saeedifar, M.A. Najafabadi, M. Fotouhi, D. Zarouchas, The study of buckling and post-buckling behavior of laminated composites consisting multiple delaminations using acoustic emission, Thin-Walled Struct. 127 (2018) 145–156, https://doi.org/10.1016/j.tws.2018.02.011. Jun.

Society, PHM, Prognostics and Health Management Society, 2020, https://doi.org/10.1109/ICPHM49022.2020.9187031. Jun.

[47] A. Ellison, H. Kim, Shadowed delamination area estimation in ultrasonic C-scans of impacted composites validated by X-ray CT, J. Compos. Mater. 54 (4) (2020) 469–482, https://doi.org/10.1177/0021998319865311.

[48] V. Giurgiutiu, Lamb wave generation with piezoelectric wafer active sensors for structural health monitoring, Proc. SPIE 5056 (2002) 17–29, https://doi.org/10.1117/12.475682. Mar.

[49] A.R. Nandyala, A.K. Darpe, S.P. Singh, Damage severity assessment in composite structures using multi-frequency lamb waves, Struct. Health Monit. 21 (6) (2022) 2391–2406, https://doi.org/10.1177/14759217221076736. Nov.

[50] F. Ricci, E. Monaco, L. Maio, N.D. Boffa, A.K. Mal, Guided waves in a stiffened composite laminate with a delamination, Struct. Health Monit. 15 (1) (2016) 3–8, https://doi.org/10.1177/1475921716636335. Jan.

[51] A.R. Diogo, B. Moreira, C.A.J. Gouveia, J.M.R.S. Tavares, A review of signal processing techniques for ultrasonic guided wave testing, Metals. (Basel) 12 (6) (2022), https://doi.org/10.3390/met12060936. Jun.

[52] M. Rautela, J. Senthilnath, E. Monaco, S. Gopalakrishnan, Delamination prediction in composite panels using unsupervised-feature learning methods with wavelet-enhanced guided wave representations, Compos. Struct. 291 (2022), https://doi.org/10.1016/j.compstruct.2022.115579. Jul.

[53] B. Zima, M. Rucka, Application of wavelet transform in analysis of guided wave propagation signals for damage detection in a steel plate, Mech. Syst. Signal. Process. 60–61 (2015) 365–378, https://doi.org/10.1016/j.ymssp.2015.02.006. Aug.

[54] H. Chen, M.J. Zuo, X. Wang, M.R. Hoseini, An adaptive Morlet wavelet filter for time-of-flight estimation in ultrasonic damage assessment, Measurement 43 (4) (2010) 570–585, https://doi.org/10.1016/j.measurement.2010.01.001. May.

[55] D. Opitz, R. Maclin, Popular ensemble methods: An empirical study, J. Artif. Intell. Res. 11 (1999) 169–198. Aug.

[56] H. Yan, W. Xie, B. Gao, F. Yang, S. Meng, A deep learning approach to impact localization and uncertainty assessment in CFRP composites using sparse PZTs: Integrating experiments and simulations, Thin-Walled Struct. 212 (2025) 113143, https://doi.org/10.1016/j.tws.2025.113143. Jul.

[57] H. Drucker, C.J.C. Burges, L. Kaufman, A.J. Smola, V. Vapnik, A support vector regression machines, Adv. Neural Inf. Process. Syst. 9 (1997) 155 161.

[58] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[59] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139, https://doi.org/10.1006/jcss.1997.1504.

[60] Compass ASTM D7137 D7137M-17, Standard Test method for compressive residual strength properties of damaged polymer matrix composite plates 1", doi:10.1520/D7137_D7137M-17.

[61] Henkel Corporation, *LOCTITE® EA 9394 AERO epoxy paste adhesive (known as Hysol® EA 9394)* [Product data sheet] (n.d.), Henkel, 2025. Retrieved April 20from, https://www.henkel-adhesives.com.

[62] PI Ceramic. (n.d.). *Piezoelectric ceramic products: Piezo technology* [Product brochure]. Retrieved April 20, 2025, from https://www.piceramic.com.

[63] D.P. Kingma and J. Ba, "Adam: a method for stochastic optimization" Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980.

[64] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting" 2014.