A novel machine learning model to design historical-independent health indicators for composite structures

Moradi, Morteza; Gul, Ferda C.; Zarouchas, Dimitrios

DOI
10.1016/j.compositesb.2024.111328

Publication date
2024

Document Version
Final published version

Published in
Composites Part B: Engineering

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

A novel machine learning model to design historical-independent health indicators for composite structures

Morteza Moradi [a,*], Ferda C. Gul [a], Dimitrios Zarouchas [a]

[a] *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

Developing comprehensive health indicators (HIs) for composite structures encompassing various damage types is challenging due to the stochastic nature of damage accumulation and uncertain events (like impact) during operation. This complexity is amplified when striving for HIs independent of historical data. This paper introduces an AI-driven approach, the Hilbert transform-convolutional neural network under a semi-supervised learning paradigm, to designing reliable HIs (fulfilling requirements, referred to as 'fitness'). It exclusively utilizes current guided wave data, eliminating the need for historical information. Ensemble learning techniques were also used to enhance HI quality while reducing deep learning randomness. The fitness equation is refined for dependable comparisons and practicality. The methodology is validated through investigations on T-single stiffener CFRP panels under compression-fatigue and dogbone CFRP specimens under tension-fatigue loadings, showing high performance of up to 93% and 81%, respectively, in prognostic criteria.

1. Introduction

Composite structures, valued for their lightweight and high-strength attributes, are increasingly utilized in industries like aerospace. Accurately predicting the health state and remaining useful life (RUL) [1,2] of these structures is crucial for safety-critical applications. However, this task is challenging due to stress redistribution from their non-homogeneous and multi-interface structure, along with potential manufacturing imperfections [3,4]. In that sense, structural health monitoring (SHM) [5] becomes essential to reveal complex historical-dependent patterns, such as progressive damage scenarios in composite structures, and unexpected occurrences, such as a bird strike on an aircraft in operation [6]. While SHM is crucial for diagnostics, prognostics and health management (PHM) [7,8] is an emerging technology as an extension of SHM that is more thorough and includes RUL prediction.

A health indicator (HI) is a valuable index that is required to first indicate the structure's health status (diagnostics) and then to predict its RUL (prognostics) [9,10]. However, measuring or estimating the health status of composite structures in a comprehensive way, where all types of damage and degradation are taken into account, is not feasible yet, especially during operation and cyclic fatigue loading. Thus, designing a

HI that meets the requirements for both diagnostics and prognostics is challenging. The requirements for a HI from the prognostics' standpoint are: If no maintenance and self-healing take place, a structure's HI should decrease throughout operational conditions due to damage accumulation. This fact should be incorporated into the design of a HI and examined using a metric known as monotonicity (Mo) [11]. The comprehensive HIs of an ensemble of associated units that have reached their end-of-life (EoL) should ideally arrive at the same value, signifying the failure threshold. However, HIs at the EoL change and do not always end up with an identical value; this discrepancy can be quantified using a metric called prognosability (Pr) [12]. HIs are more predictable if they have comparable trends and a similar correlation in terms of usage time for similar units. By using the trendability (Tr) criterion [9,12], it is possible to quantify the resemblance between HIs. A HI must satisfy the three evaluation criteria of Mo, Pr, and Tr from the viewpoint of prognostics, which is the primary focus of the current work.

The fact that RUL prediction and HI construction models are historical-dependent is a common drawback. It means that to enhance the performance of the HI and RUL prediction models, the temporal relationship between historical data from the starting point until the present moment should be considered [1,13,14]. As a result, prognostic and HI construction models function less efficiently when prior information, either entirely or partially from the beginning, is missing. This

Abbreviations

AI	Artificial Intelligence
BiLSTM	Bidirectional LSTM
C–C	Compression-Compression
CFRP	Carbon Fiber Reinforced Polymer
CNN	Convolutional Neural Network
D	Dropout
DL	Deep Learning
EL	Ensemble Learning
EoL	End-of-Life
FC	Fully Connected
FIR	Finite Impulse Response
GW	Guided Wave
GWSHM	Guided Wave-based Structural Health Monitoring
HI	Health Indicator
HT	Hilbert Transform
HT-FIR	Hilbert Transform FIR
HT-SSCNN	Hilbert Transform - Semi-Supervised CNN
Leaky ReLU	Leaky Rectified Linear Unit
LOOCV	Leave-One-Out Cross-Validation
LSTM	Long Short-Term Memory
MMK	Modified Mann-Kendall
Mo	Monotonicity
MSE	Mean Squared Error
PHM	Prognostics and Health Management
Pr	Prognosability
RMSE	Root-Mean-Square Error
RUL	Remaining Useful Life
SAE	Simple Averaging Ensemble
SHM	Structural Health Monitoring
SSCNN	Semi-Supervised CNN
SSL	Semi-Supervised Learning
SSP	Single-Stiffened Composite Panel

T-T	Tension-Tension
Tr	Trendability
WAE	Weighted Averaging Ensemble

Symbols

$x(t_p)$	HI at the time of t_p
$x(t_i)$	HI at the time of t_i
M	number of units (specimens)
x_j	HI for the jth unit out of M units
x_k	HI for the kth unit out of M units
N_j	Number of measurements for the jth unit
N_k	Number of measurements for the kth unit
$cov(x_j, x_k)$	Covariance where x_j and x_k are HIs
$sgn(\bullet)$	Signum function
σ_{x_j}	Standard deviation of x_j
σ_{x_k}	Standard deviation of x_k
τ	Test unit
τ'	All units (except for the test unit)
M^τ	Number of test units
$M^{\tau'}$	Number of all units
T_i	Simulated HI as target value at time step i
HI_i	Network output for time step i
$\mathscr{H}\{\bullet\}$	Hilbert transform
j	Imaginary part
β	Shape parameter
t	Operating time
N_a	Number of actuators
N_s	Number of sensors
f_k	k^{th} base learner model
ω_k	weight for k^{th} base learner model
$\overline{\omega}_k$	normalized weight for k^{th} base learner model
$L_{regress}$	Loss function of regression model

might occur for various reasons during operation due to unavailable monitoring systems, misfunctioned sensors, etc. This paper addresses the aforementioned challenge by developing a new data-driven approach designed with signal processing and artificial intelligence (AI) elements.

Firstly, a robust SHM method is needed to extract informative, historical-independent patterns. The guided wave-based SHM (GWSHM) technique can be a potential candidate that can generate such historical-independent patterns [15]. This active SHM technique exhibits the capability to estimate the size and location of damage that has progressed from the initial stages up to the current time [16,17]. It can also be used to estimate the stiffness reduction in the structure resulting from progressive damage and material degradation, based on only the most recent recorded data [18]. GWs are among the most widely utilized SHM methods for thin-walled structures in the aviation industry owing to its capability of interrogating the whole structure with a low attenuation ratio and detecting small-sized damages with high accuracy [19, 20]. However, GWs pose a challenge as they not only convey information about damage but also carry signals that are susceptible to interference from reflection and scattering phenomena at structural boundaries. Environmental conditions can further affect the reliability of GW signals [21,22]. Moreover, GWs are dispersive and have a multimodal nature, meaning that many wave modes may exist in a recorded signal, and their characterization depends on the frequency, material stiffness, density, and thickness of the structure [23]. Different structural geometries under diverse loading conditions necessitate varying numbers and positioning of sensors, often requiring distinct data processing algorithms. Therefore, the diagnosis and prognosis of

composite structures via GWs is a hard task when it comes to correlating the GW signals with the health state of the entire structure. This paper introduces an adaptive approach designed to flexibly accommodate different GW sensor numbers, networks, and setups to address variations in the geometry, layups, and loading conditions of composite structures.

GW signals have been investigated to discriminate specific damage modes, such as delamination and matrix cracks [24], or to predict damage severity under fatigue loading [25,26]. However, since the degradation of composite structures is a combined process of multiple damage mechanisms evolving stochastically, the quantification of a specific damage type and location may create limited information for the health state projection of the structure. A HI may be seen as adopting engineers' perspectives and attempting to determine how all the distinct (microcrack, crack, delamination, fiber breakage, etc.) and spatially dispersed fatigue damage contributes to overall structural deterioration, which will be useful from a PHM and structural design standpoint.

In addition to the previously discussed aspects, it is important to note that translating GW data to the appropriate HI value at each time step when GW inspections are conducted is a challenging task. In fact, a model is needed to map thousands of data points (as can be seen in experimental campaigns that generated extensive datasets—cite NASA and ReMAP) to a single HI value at the current time, regardless of the prior HIs. To address this challenge, data-driven approaches, especially AI, have drawn attention in diagnostic [27–30] and prognostic [13,31] applications thanks to their ability to discover complex and nonlinear relationships between data. Nevertheless, several constraints exist to extract proper HIs from GW data:

1. Composite structures lack comprehensive HI labels, limiting the applicability of supervised models for this regression task.
2. The data used for modeling is only available at individual GW inspection time steps, without access to prior or posterior data. However, the HI model requires optimization considering the entire historical trajectory, posing challenges in efficiently updating model parameters; e.g., an error value is needed to tune the weights of an artificial neural network (ANN) for each backward followed by the feedforward, which in this case is not doable directly based on evaluation criteria.
3. Although the set of evaluation criteria (i.e., Mo, Pr, and Tr) can be implemented into an objective function termed the "fitness" function, it is not differentiable [31]. Thus, fitness cannot be implemented in the loss function of back-propagation-based algorithms like ANN. While derivative-free models like Bayesian optimization and reinforcement learning exist, they face constraints in handling a high number of inputs, require feedback from the objective function for each iteration, and are comparatively slower than back-propagation-based models.

To tackle the above-mentioned challenges as well as previous limitations, an approach called Hilbert transform - semi-supervised convolutional neural network (HT-SSCNN) is introduced to construct HIs fulfilling evaluation criteria based on historical-independent GW data. First, the GW signals recorded from the network of sensors and their envelopes are extracted using the Hilbert transform (HT). Then, these envelopes are integrated to reshape a 3D form input to feed the subsequent deep learning (DL) model. Convolutional neural network (CNN) architectures are designed to fuse the 3D inputs trained by a semi-supervised learning (SSL) paradigm, and their inherent randomness and uncertainty are mitigated by ensemble learning (EL). All GW data generated from different triggering frequencies is also fused during the EL step. Six methods, including four average-based and two DL networks, are considered for EL.

Two main datasets are investigated to validate and discuss the proposed methodology.

1. ReMAP[1]: compression-compression (C–C) fatigue tests on T-single stiffener composite panels, which include impacts and disbond damage [32].
2. NASA: tension-tension (T-T) fatigue tests on composite dogbone specimens with different layups [33].

The fitness function is redefined for evaluating the test phase of the DL model to provide a more trustworthy foundation for comparison and enhance the practical reliability of the standard. Ablation experiments are conducted, including variations in dataset division and leave-one-out cross-validation (LOOCV), confirming the generalizability of the approach. The contribution of this work lies in the development of a robust and reliable HI construction model, generating historical-independent HIs suitable for RUL estimation of composite structures with different layups, geometry, and loading conditions.

2. Background

Two confirmed evaluation metrics of HIs in the field of PHM are employed to evaluate the proposed methodology. The first metric is the *Fitness* function, compounding prognostic criteria as described in Section "HIs' criteria." The root-mean-square error (RMSE) is another metric to measure deviation from the ideal HIs.

2.1. HIs' criteria

This subsection begins by introducing the prognostic metrics, which will be modified in the present work for consideration of only test units rather than both training and test units.

I. HIs' criteria on all units

Due to the high importance of HIs' evaluation metrics amongst different stages of the methodology, they are introduced at the outset. The evaluation criteria of HIs is based upon three established criteria, namely Mo, Pr, and Tr [31,34]. These metrics are defined as follows:

$$Mo = \frac{1}{M} \sum_{j=1}^{M} \left| \frac{1}{N_j - 1} \sum_{i=1}^{N_j} \frac{\sum_{p=1,p>i}^{N_j} (t_p - t_i).sgn(x(t_p) - x(t_i))}{\sum_{p=1,p>i}^{N_j} (t_p - t_i)} \right| .100\% \qquad (1)$$

$$Pr = exp\left(-\frac{\sqrt{\frac{1}{M}\sum_{j=1}^{M} \left| x_j(N_j) - \left[\frac{1}{M}\sum_{i=1}^{M} x_i(N_i) \right] \right|^2}}{\frac{1}{M}\sum_{j=1}^{M} |x_j(1) - x_j(N_j)|} \right) \qquad (2)$$

$$Tr = \min_{j,k} \left| \frac{cov(x_j, x_k)}{\sigma_{x_j} \sigma_{x_k}} \right|; \quad j,k = 1,2,\ldots,M \qquad (3)$$

where $x(t_p)$ and $x(t_i)$ indicate HI at the times of t_p and t_i, respectively. The $sgn(\bullet)$ is the signum function and the $cov(x_j, x_k)$ is the covariance, where x_j and x_k are HIs for the jth and k^{th} unit (out of M units – composite specimens) with N_j and N_k measurements (time steps), respectively. σ_{x_j} and σ_{x_k} denote standard deviations of x_j and x_k, respectively. Modified Mann-Kendall (MMK) is the evaluation metric chosen for Mo in Eq. (1). MMK is more noise-resistant than the Sign and Mann-Kendall versions and takes into account the correlation between data points with time gaps greater than one time unit [31,35]. Ratings for the three HIs metrics (Mo, Pr, and Tr) range from 0 to 1, with a score of 1 denoting ideal HI performance. In light of these criteria, the following formulation for the fitness metric is used:

$$Fitness = a.Mo_{HI} + b.Pr_{HI} + c.Tr_{HI} \qquad (4)$$

The fitness metric for HIs ranges from 0 (indicating minimal quality) to 3 (indicating maximal quality), assuming that the control constants a, b, and c are each equal to 1. The subscript HI refers to "health indicator".

II. HIs' criteria on test units

The aforementioned HIs criteria have been formulated to apply to all units (composite coupons) under monitoring, specifically from their initial to final failure status. With this in mind, the evaluation of HIs' quality lacks implication without access to complete trajectories of HIs across all units. Therefore, the inclusion of all units becomes crucial to accurately rating fitness, whether during the training or testing phase of data-driven models. However, a potential issue arises due to the possibility of highly matched HIs during the training phase, which could result in a misleadingly high fitness score when confronted with an unmatched HI from a specific unit during the testing phase. Thus, HI metrics must also be assessed in relation to only the test unit. The singular noun "test unit" was chosen over the plural "test units" to better reflect the unique context of this work, while it could be easily expanded to accommodate multiple test units.

The HIs criteria, in particular Pr, are redefined in this study to concentrate more on the test units than the entire set of units (train/validation/test). These updated metrics aim to assess whether the test unit significantly deviates from the training units and, if so, to what
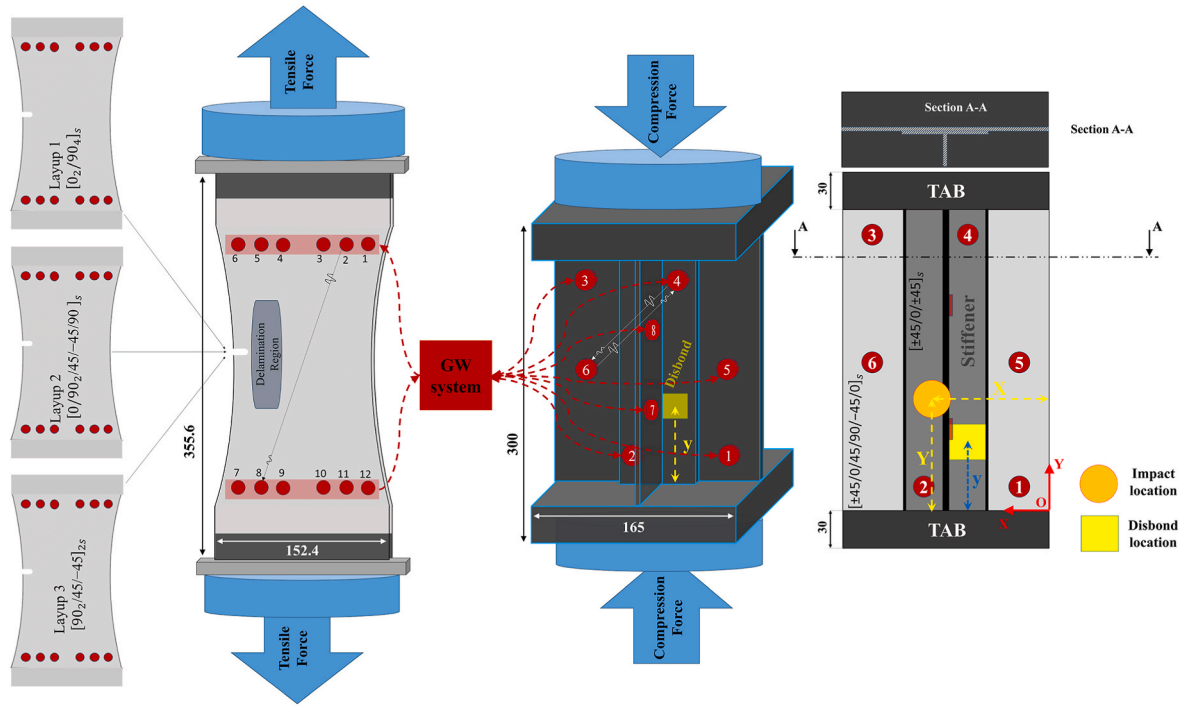

Fig. 1. Composite dogbone specimens under T-T fatigue (left) and single T-stiffener CFRP panel under C–C fatigue (right) being monitored by PZT sensors (red circles) (dimensions in [mm]). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

extent. It is important to note that the methodology may not work if the metrics produce higher values only for the test unit while remaining lower for the training units. On the other hand, attaining high scores for the test unit becomes significant if the metrics are consistently high for the training units. For instance, consider the Mo metric, which may exhibit a high value only for test units, not across all units, due to non-monotonic HIs in the training units. In this situation, the prognostic model may struggle to predict RUL given such HIs because it fails to learn the proper monotonic pattern from the training units, even if the test unit displays a monotonic behavior. Another example is illustrated by the Pr metric, where the Pr score for a test unit might be high due to its low deviation from the average of the training HIs at EoL (deviation basis). Simultaneously, the standard deviation of HIs at EoL for training units could be high, resulting in a lower Pr score when considering all units. Once again, a prognostic model may face challenges in predicting RUL when trained with non-prognostable HIs of training units exhibiting high variance at EoL.

The internal summation of Eq. (1) yields the same definition of Mo for a single unit:

$$Mo^{\tau} = \left| \frac{1}{N_{\tau}-1} \sum_{i=1}^{N_{\tau}} \frac{\sum_{p=1,p>i}^{N_{\tau}} (t_p - t_i).sgn(x(t_p) - x(t_i))}{\sum_{p=1,p>i}^{N_{\tau}} (t_p - t_i)} \right| .100\% \qquad (5)$$

where the symbol τ denotes the test unit. When more than one unit is being considered for testing, Eq. (5) will be similar to Eq. (1), with the averaging over only the test units. However, Pr needs to be redefined. In this adjustment, the numerator is computed using the deviation of the HI at EoL for the test unit from the deviation basis (i.e., its corresponding value averaged over the training units), rather than the standard deviation of HIs at EoL across all units:

where τ' denotes all units set except for the test ones and $M^{\tau'}$ denotes the number of those units (including the training units and even validation units). It should be noted that $x_j(1)$ and $x_j(N_j)$ denote the HI values of the jth unit at the initiation and EoL, respectively. As a scaling factor, the denominator in this case represents the mean value of the difference between HIs at the outset and the EoL across all units ($\tau \cup \tau'$) or the training units (τ'). After examining both options, the mean value over all units is adopted (as symbolized in Eq. (6)). When dealing with multiple

$$Pr^{\tau} = exp\left(-\frac{\left| x_{\tau}(N_{\tau}) - \overbrace{\left[\frac{1}{M^{\tau'}} \sum_{i=1}^{M^{\tau'}} x_i(N_i) \right]}^{\text{deviation basis}} \right|}{\underbrace{\frac{1}{M} \sum_{j=1}^{M} \left| x_j(1) - x_j(N_j) \right|}_{\text{scaling factor}}} \right) = exp\left(-\frac{\left| x_{\tau}(N_{\tau}) - \overbrace{\mu_{M^{\tau'}}}^{\text{deviation basis}} \right|}{\underbrace{\sigma_M}_{\text{scaling factor}}} \right) \qquad (6)$$

Table 1
Information of single T-stiffener CFRP panels tested under C–C fatigue loading (ReMAP dataset). Dimensions are in [mm].

Name	L1-03	L1-04	L1-05	L1-09	L1-23
Name index	SSP1	SSP2	SSP3	SSP4	SSP5
X-location of impact	50	25	115	82.5	–
Y-location of impact	80	80	160	140	–
Time of Impact	at 0 cycles		at 0 cycles	at 0 cycles	–
Size of disbond	–	–	–	–	30 × 30
y-location of disbond	–	–	–	–	60
Min Load (kN)	−6.5	−6.5	−6.5	−6.5	−5 and −6
Max Load (kN)	−65	−65	−65	−65	−50 and −60
Cycles	152,458	280,098	144,969	133,281	438,000

units (M^τ) for testing, the EoL deviation basis can be established from the training set (τ'), the test set (τ), or a combination of both ($\tau \cup \tau'$), with the latter being advisable for M units.

Regarding Tr, the minimum correlation of HIs should be calculated between distinct units, which is challenging when assessing only one test unit while excluding training ones. If the correlation between the HIs of the test and training units is computed pairwise and the minimum value is selected as Tr, it might exceed the correlation calculated given all units' HIs pairwise. To ensure a more strict evaluation, the same formula as Eq. (3) is applied to the test units. With these updated metrics, the fitness metric for the test unit is:

$$Fitness^\tau = a.Mo_{HI}^\tau + b.Pr_{HI}^\tau + c.Tr_{HI} \tag{7}$$

where τ indicates "HIs' criteria on test units".

2.2. Deviation from the simulated ideal HIs

The second evaluation criterion is RMSE, calculated as:

$$RMSE = \sqrt{\frac{1}{N_j} \sum_{i=1}^{N_j} (T_i - HI_i)^2} \; ; \quad j \in M \tag{8}$$

In this formula, N_j is the length of the sequence (*HI*), representing the number of data points (time steps of GW inspection). T_i represents the target value (simulated HI under SSL paradigm) and HI_i is the network's output for time step *i*. The RMSE provides a single score for each unit's constructed HI, with a primary focus on reporting the test unit's HI score, excluding training or validation data.

3. Experimental campaigns

In this study, two different datasets have been investigated and analyzed separately. The first dataset, ReMAP, contains five composite skin-stiffener panels that were subjected to C–C fatigue loading using an MTS machine with a frequency of 2 Hz and an R-ratio of 10. These panels consisted of a skin panel and a T-stiffener based on an Embraer design. Both the skin and stiffener were constructed using IM7/8552 carbon fiber-reinforced epoxy unidirectional prepreg, with specific layups of

$[\pm 45/0/45/90/\text{-}45/0]_S$ and $[\pm 45/0/\pm 45]_S$, respectively. Impact loading of 10 J was applied at various locations in the stiffener region (see Fig. 1 and Table 1), and one panel also had disbond defects intentionally introduced between the skin and T-stiffener during manufacturing (additional information can be found in Table 1). In this experiment, various techniques were employed to monitor the composite panels [32], including the GW method with a 5000-cycle interval, which is the focus of this research. GW data collection is carried out with eight surface-attached PZT sensors, where four sensors are located on the skin panel, two sensors are located on top of the stiffener-skin bondline, and two sensors are attached on the stiffener web. The GW system operates with one PZT serving as the actuator with six excitation frequencies (50 kHz, 100 kHz, 125 kHz, 150 kHz, 200 kHz, and 250 kHz) and the remaining seven PZTs functioning as sensors, rotating through all eight PZTs.

The second dataset, NASA, has three different layups, Layup 1, Layup 2, Layup 3, with the ply orientation, $[0_2/90_4]_S$, $[0/90_2/\pm 45/90]_S$, and $[90_2/\pm 45]_{2S}$, respectively [33,36,37]. Torayca T700G unidirectional carbon-prepreg material has been used to manufacture the dogbone geometry coupons. The coupons with a notch have been submitted to T-T fatigue load, and the damage accumulation has been quantified through X-ray imaging for specific damage types, which are matrix cracks and delamination mainly. GW acquisition has been performed through a surface-attached PZT network with three different boundary conditions: traction-free, clamped, and loaded, among which the clamped boundary condition was considered due to its closer resemblance to real-world scenarios. The network contains one actuator and one sensor array with six linearly distributed PZT transducers in each, as shown in Fig. 1. GW acquisition is realized between the linear arrays, which makes 36 sensor paths in total. Seven excitation frequencies have been applied in the range of 150–450 kHz with 50 kHz incrementation by a 5-cycle Hanning modulated tone-burst signal which is a short-duration signal that consists of a few cycles of a sinusoidal waveform and it is often used in applications like ultrasonic testing and medical imaging. The choice of the number of cycles depends on the application's specific requirements as the short duration helps in achieving good resolution in time-domain measurements [38]. Information regarding the mechanical test parameters can be found in Table 2. Further information regarding the mechanical experiment, GW data acquisition, and X-ray imaging process can be found in Refs. [33,36,37].

4. Methodology

In this section, the methodologies and step-by-step process are introduced as can be seen in Fig. 2.

4.1. Signal processing and 3D input preparation

Prior to using DL networks, signal processing techniques with explicit and fast solutions can improve performance and simplify following DL modeling. An effective method involves extracting GW signal envelopes using the magnitude of their analytic signals. This is achieved through the Hilbert transform (HT):

$$Z(t) = x(t) + j\mathscr{H}\{x(t)\} \quad ; \quad \mathscr{H}\{x(t)\} = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t - \tau} d\tau \tag{9}$$

Table 2
Information of the dogbone CFRP plates tested under T-T fatigue loading (NASA dataset).

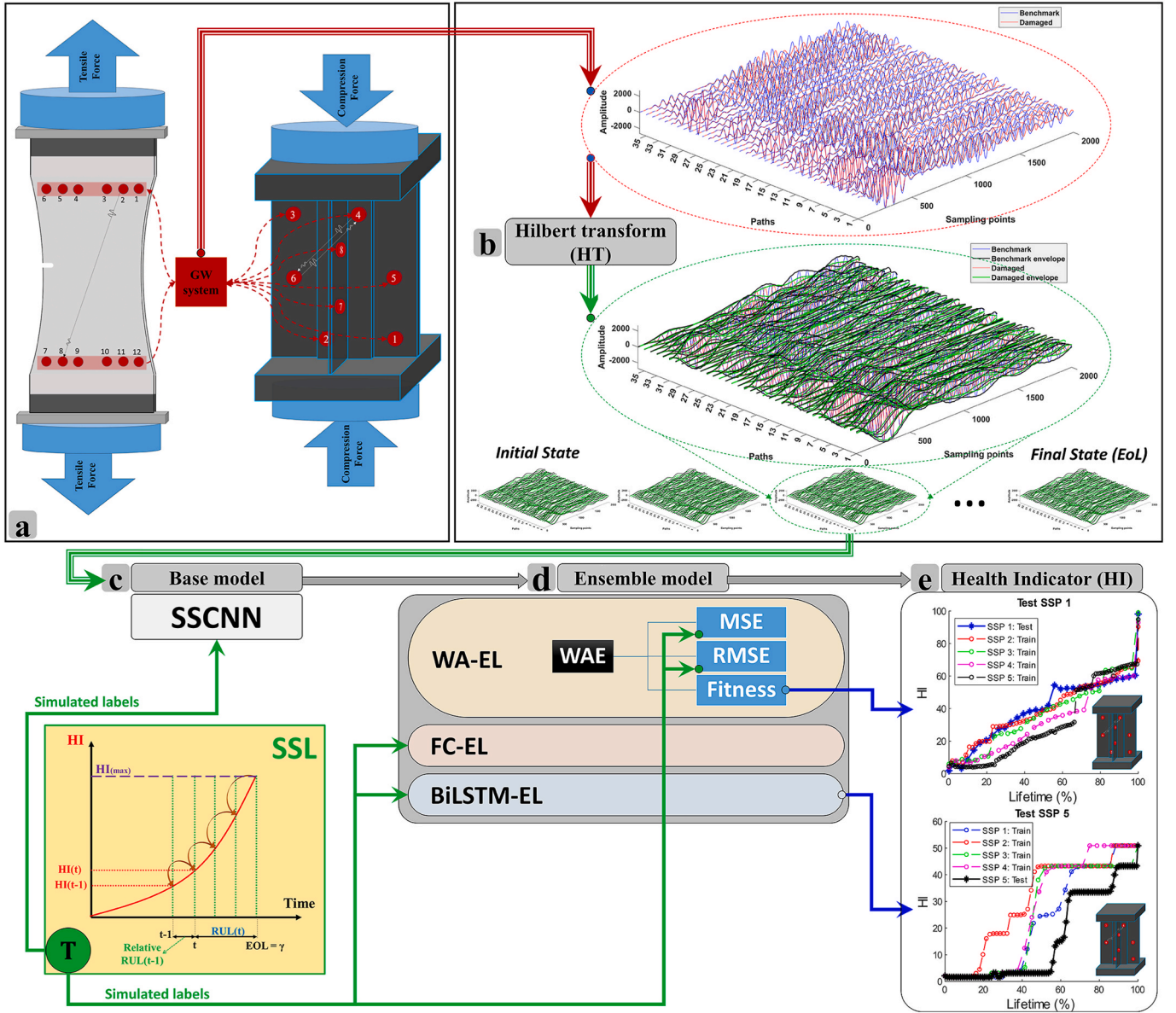| Ply orientation	Layup 1: $[0_2/90_4]_S$				Layup 2: $[0/90_2/\pm 45/90]_S$			Layup 3: $[90_2/\pm 45]_{2S}$			
Name index	S1	S2	S3	S4	S1	S2	S3	S1	S2	S3	S4
Static Failure (Mpa)	133.8				147.5			104.8			
Max Load (kN)	111.7				127.5			89.6			
Load Ratio	0.14				0.14			0.13			
Cycles (×1000)	*227*	100	1650	125	207	900	1250	150	300	895	750

Fig. 2. The overall proposed framework: (a) guided wave (GW) monitoring; (b) signal processing (Hilbert transform); (c) base learner model (SSCNN); (d) ensemble learner model; (e) extracted health indicator (only shown for single T-stiffener CFRP panels).

where $x(t)$ is the input signal, j is the imaginary unit, and $\mathcal{H}\{\bullet\}$ is the Hilbert transform. This is replaced by a finite impulse response (FIR) filter in discrete-time signal processing to reduce computational complexity [39]. This specific FIR filter is termed the Hilbert transform FIR (HT-FIR) filter, with its length determined by the excitation frequency in the current work. For example, a 400 kHz frequency corresponds to an HT-FIR filter length of 400. The filter is created by applying a Kaiser window to minimize the effects of spectral leakage and to control the shape of the frequency response of the signal. The Kaiser window is characterized by a parameter, often denoted as β, that allows for adjusting the trade-off between the main lobe width and the side lobe levels [40,41]. The shape parameter selected as β = 8 to an ideal brick-wall filter. Similar processing generates upper and lower envelopes for all GW signals.

A 3D form of data (according to [paths between actuators and sensors] × [signal length] × [upper and lower envelopes in two states]) is implemented to prepare inputs for the subsequent SSCNN model. Accordingly, the input dimensions are 36 × 2000 × 4 and 56 × 2000 × 4 for the NASA and ReMAP datasets, respectively. Fig. 3 shows, for the

NASA dataset (Layup 1), the received GW signals excited by one frequency (150 kHz), the extracted envelopes for all paths, and the pertinent 3D input of SSCNN at cycle 60000.

4.2. Base learner model

This section introduces the learning paradigm, the DL architecture, and the training adjustments.

Learning Framework: In the absence of true HI values, an idealized function is used to generate HI labels, and an SSL paradigm is employed by incorporating prognostic metrics (Mo, Pr, and Tr) while leveraging existing EoL [31,42]. This approach belongs to the category of intrinsically semi-supervised inductive learning algorithms, which extend traditional supervised methods to use both labeled and unlabeled data for optimizing an objective function. The optimal generator function is represented as a quadratic polynomial:

$$HI_{(t)} = \frac{t^2}{t_{EoL}^2} \tag{10}$$
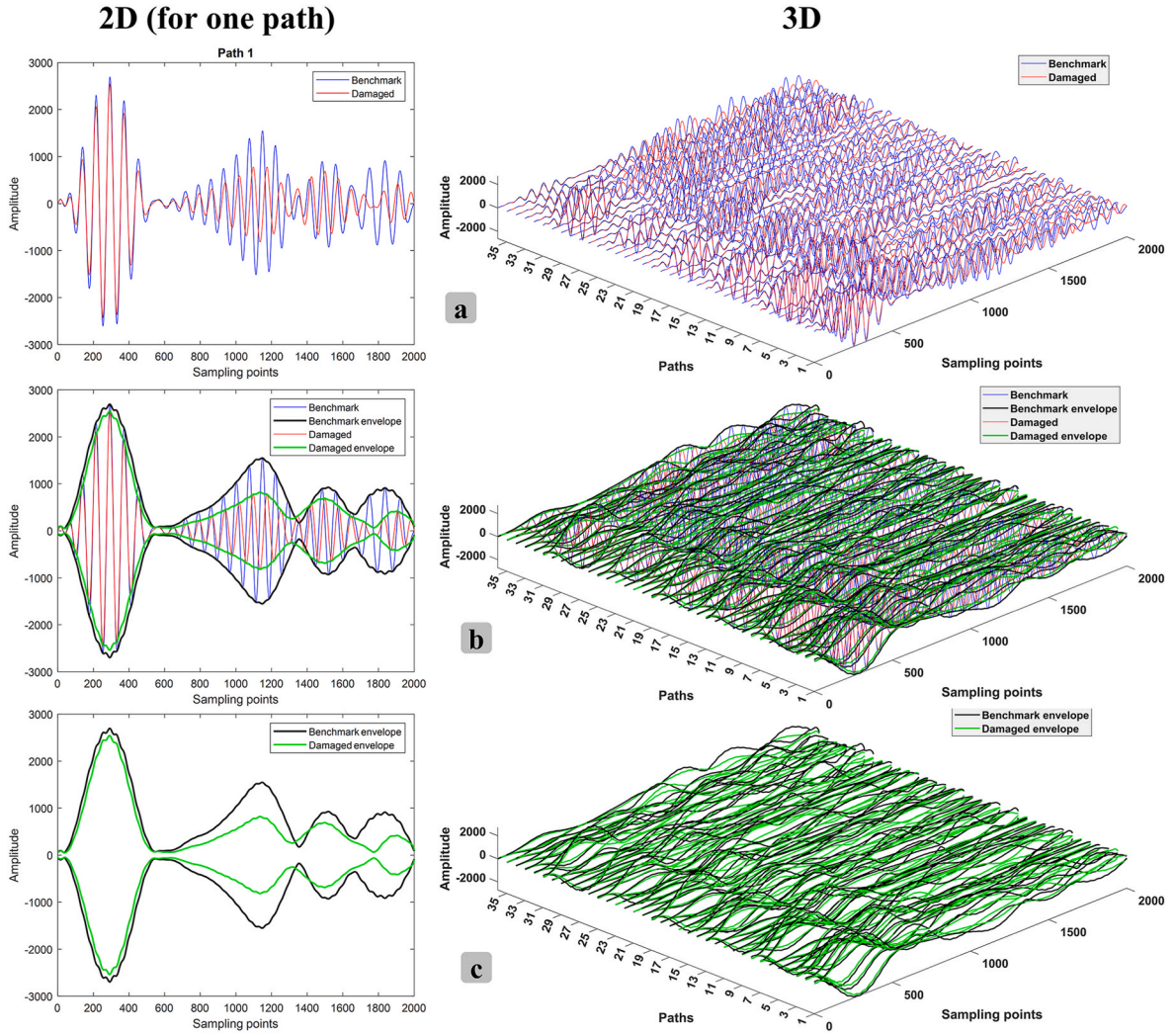
2D (for one path)   3D

Fig. 3. Right column: (a) Sensed GW signals excited by 150 kHz and (b) their envelopes for all 36 paths of the NASA dataset (Layup 1) at baseline and cycle 60000, as well as the relevant (c) 3D input of SSCNN at cycle 60000. Left column: Their 2D display for only one path.

where t is the operating time and t_{EoL} is the time of the final failure. Eq. (10) results a HI in the range [0, 1], which can also be scaled by a multiplication in a coefficient [13,43]. The normalization via t_{EOL}^2 is necessary to obtain the desired values for Pr since each unit has a different lifetime. The labels in this study are scaled by a factor of 100, resulting in a range from 0 (healthy state) to 100 (failure threshold, EoL). It is important to note that this hypothetical HI is only utilized for the available training samples where lifetimes are known in advance.

DL architecture and adjustments: A CNN architecture, illustrated in Fig. 4, has been designed to map GW inputs to the simulated ideal HI. The inputs, as previously described, take the 3D shape of $(N_a \times N_s) \times 2000 \times 4$, where N_a and N_s represent the number of actuators and sensors, respectively. The Leaky Rectified Linear Unit (Leaky ReLU) is a variation of the ReLU activation function, distinguished by a small slope for negative values. In contrast to the flat slope of ReLU, Leaky ReLU introduces a small, positive gradient for inactive units, addressing the vanishing gradient problem [44]. The leakage coefficient, the small slope, for all Leaky ReLU functions is set at 0.01, after trial and error. To calculate the loss function between predictions and targets, a mean squared error (MSE) is employed:

$$L_{regress} = \frac{1}{R} \sum_{i=1}^{R} (T_i - HI_i)^2 \tag{11}$$

where $T_i = t_i^2 / t_{EOL}^2$ denotes the target value and HI_i is the output of the last layer. An Adam optimizer [45], a stochastic gradient descent method that is based on adaptive estimation of first-order and second-order moments, was utilized to train the SSCNN with 38400 and 28800 learnable parameters for ReMAP and NASA setups. After trial and error, an initial learning rate of 0.001 was employed.

The training time for the SSCNN model on NASA and ReMAP datasets is 161 s and 571 s, respectively, using a laptop with an Intel Core i7-8665U CPU and 16 GB RAM for 200 epochs with a batch size of 10. The variation in training time between datasets is attributable to differences in input dimensions (depending on the number of paths between actuators and sensors—36 for NASA and 56 for ReMAP) as well as the number of training units and GW inspections (38 GW inspections for 3 units of NASA (Layup 1) and 120 GW inspections for 4 units of ReMAP). To isolate the influence of the number of training units and GW inspections, the time required to train the SSCNN model for one batch with a size of 10 and one epoch is 0.201 s and 0.238 s, given 36 to 56 paths as input, respectively.

4.3. Ensemble learner models

Ensemble learning (EL) techniques mitigate randomness and uncertainties in ML models post-HI extraction. They come in three main categories: bagging [46], boosting [47], and stacking [48]. From
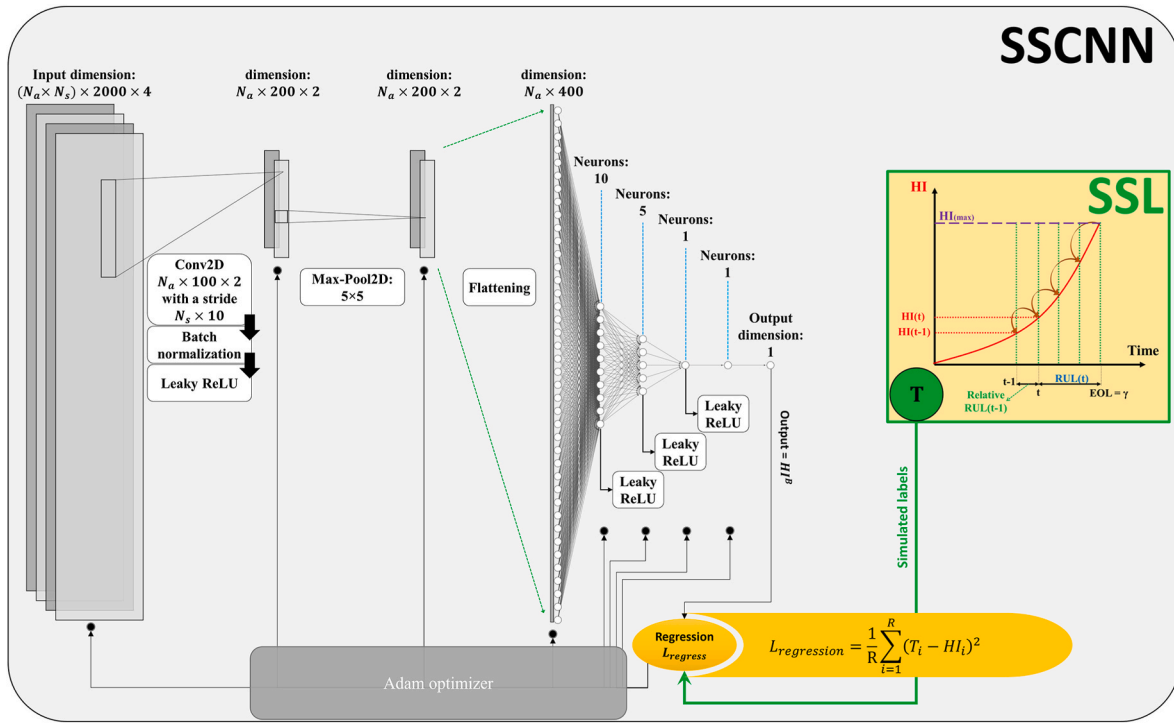
Fig. 4. The architecture of the semi-supervised convolutional neural network as the base learner.

Table 3
Ensemble learner models.

Model num.	Model name	Architecture (hidden layers)						
1	**SAE**							
2	**WAE-MSE**							
3	**WAE-RMSE**							
4	**WAE-Fitness**							
5	**FC-Net**	FC (100)	D (0.5)	ReLU	FC (5)	D (0.5)	ReLU	FC (1)
6	**BiLSTM-Net**	D (0.5)	BiLSTM (5)	D (0.5)	FC (5)	D (0.5)	ReLU	FC (1)

another point of view, EL models that rely on averaging can be mainly divided into two types: simple averaging ensemble (SAE) [49] and weighted averaging ensemble (WAE) [50], with the first being a case of the last. This work prioritizes averaging ensemble models with various weighting techniques and DL ensemble models due to their superior performance.

The initial step involves leave-one-out cross-validation (LOOCV) [51]. Here, one unit (composite specimen) is set aside for testing,
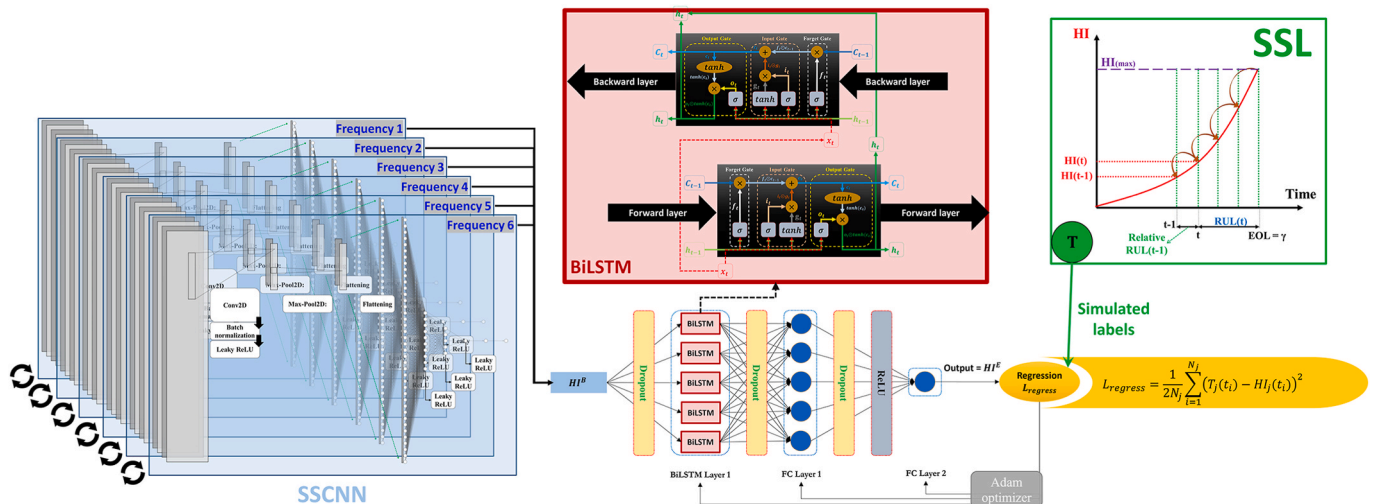

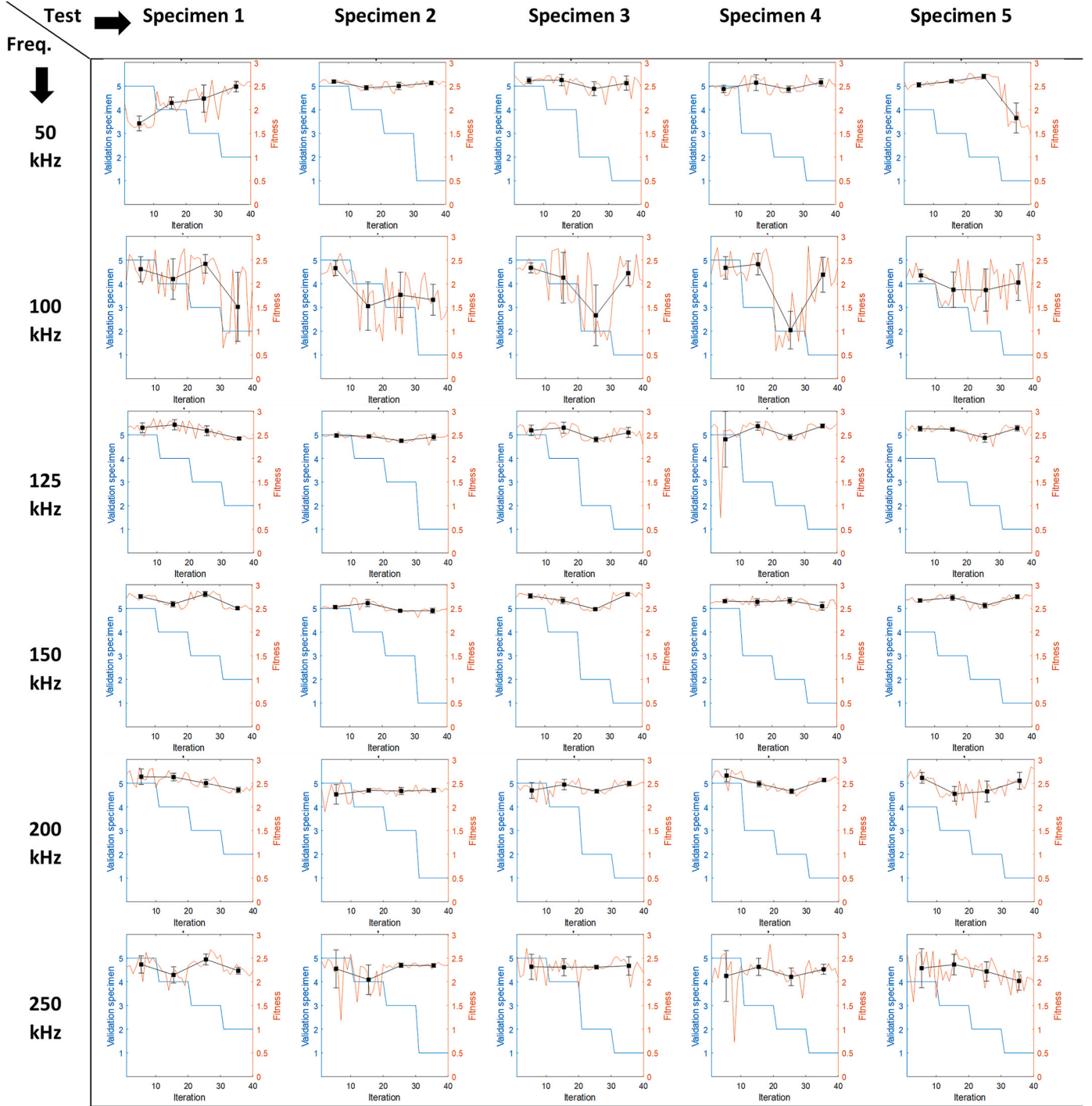**Fig. 5.** The architecture of the BiLSTM-Net ensemble learner model.

Fig. 6. Fitness for HT-SSCNN over various subsets (train, validation, and test combinations) and frequencies in the ReMAP dataset (fitness calculated from Eq. (4), considering all units referred to as 'F-All'). The blue line indicates the chosen unit for validation, the brown line represents the fitness score, and the black line illustrates the variance (Std) bars of the fitness with filled square markers denoting its mean value, plotted at every 10 iterations. Similar figures for the NASA dataset can be found in the Supplementary Materials.(For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

another for validation, and the rest are used for training. The validation unit rotates through all available options, and the base learner model (SSCNN) is trained ten times with different random seed numbers [52] for weight and bias initialization. Subsequently, the 100 HIs predicted by SSCNN are combined through a process that includes SAE, WAE, and DL models. The WAE is expressed as follows:

$$f_{WAE} = \sum_{k=1}^{K} \overline{\omega}_k f_k \quad ; \quad \overline{\omega}_k = \frac{\omega_k}{\sum_{k=1}^{K} \omega_k} \tag{12}$$

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 these weights can be determined based on various error metrics. In this

Table 4
Fitness values for base learner models averaged over the 40 repetitions for ReMAP dataset.

Frequency	Fold 1		Fold 2		Fold 3		Fold 4		Fold 5	
	F-All	F-Test	F-All	F-Test	F-All	F-Test	F-All	F-Test	F-All	F-Test
50 kHz	2.23 ($\pm$0.15)	1.55 ($\pm$0.34)	2.31 ($\pm$0.37)	2.12 ($\pm$0.45)	2.46 ($\pm$0.25)	2.22 ($\pm$0.48)	2.41 ($\pm$0.42)	2.17 ($\pm$0.35)	2.49 ($\pm$0.5)	2.28 ($\pm$0.5)
100 kHz	2.44 ($\pm$0.35)	2.25 ($\pm$0.49)	2.03 ($\pm$0.26)	1.55 ($\pm$0.3)	2.5 ($\pm$0.23)	2.23 ($\pm$0.23)	2.57 ($\pm$0.23)	2.41 ($\pm$0.38)	2.28 ($\pm$0.33)	2 ($\pm$0.37)
125 kHz	2.54 ($\pm$0.31)	2.4 ($\pm$0.44)	2.21 ($\pm$0.29)	1.93 ($\pm$0.24)	2.52 ($\pm$0.21)	2.11 ($\pm$0.35)	2.55 ($\pm$0.33)	2.36 ($\pm$0.43)	2.55 ($\pm$0.2)	2.21 ($\pm$0.31)
150 kHz	2.59 ($\pm$0.45)	2.43 ($\pm$0.48)	2.19 ($\pm$0.34)	2.01 ($\pm$0.44)	**2.72 ($\pm$0.14)**	**2.43 ($\pm$0.3)**	2.54 ($\pm$0.19)	2.14 ($\pm$0.33)	2.6 ($\pm$0.22)	2.3 ($\pm$0.29)
200 kHz	2.45 ($\pm$0.3)	2.21 ($\pm$0.45)	2.16 ($\pm$0.31)	1.91 ($\pm$0.33)	2.5 ($\pm$0.09)	2.09 ($\pm$0.24)	2.63 ($\pm$0.14)	2.44 ($\pm$0.46)	**2.62 ($\pm$0.32)**	**2.55 ($\pm$0.39)**
250 kHz	2.51 ($\pm$0.51)	2.38 ($\pm$0.52)	2.21 ($\pm$0.26)	1.93 ($\pm$0.24)	2.35 ($\pm$0.2)	1.96 ($\pm$0.35)	2.29 ($\pm$0.4)	2.09 ($\pm$0.55)	2.45 ($\pm$0.25)	1.89 ($\pm$0.21)
All	2.55 ($\pm$0.32)	2.4 ($\pm$0.46)	2.19 ($\pm$0.37)	1.84 ($\pm$0.38)	2.62 ($\pm$0.2)	2.42 ($\pm$0.3)	2.64 ($\pm$0.25)	2.48 ($\pm$0.37)	2.53 ($\pm$0.49)	2.33 ($\pm$0.55)

study, the error metrics considered are MSE, RMSE, and fitness. MSE and RMSE are computed by comparing the predictions of the kth base model ($HI^{k(E)}$, where (E) denotes the ensemble HI) and the simulated HIs (T):

$$\omega_k^{MSE} = \frac{1}{MSE\left(T, HI^{k(E)}\right)} = \frac{1}{\frac{1}{M^\tau}\sum_{j\in\tau}\left[\frac{1}{N_j}\sum_{i=1}^{N_j}\left(T_j(t_i) - HI_j^{k(E)}(t_i)\right)^2\right]} \quad (13)$$

$$\omega_k^{RMSE} = \frac{1}{RMSE\left(T, HI^{k(E)}\right)} = \frac{1}{\frac{1}{M^\tau}\sum_{j\in\tau}\sqrt{\left[\frac{1}{N_j}\sum_{i=1}^{N_j}\left(T_j(t_i) - HI_j^{k(E)}(t_i)\right)^2\right]}} \quad (14)$$

$$\omega_k^{Fitness} = a.Mo_{HI} + b.Pr_{HI} + c.Tr_{HI} ; \quad j \in \tau^{'} \quad (15)$$

When all $\overline{\omega}_k$ are set to one, the SAE approach is implemented.

In addition to averaging EL models, an alternative approach is to use a subsequent ML-based model to fuse predictions, aiming to reduce inherent randomness in base learner models. This ML-based ensemble learning model can be implemented using DL networks. Different networks with various layer types, including fully connected (FC), long short-term memory (LSTM), and bidirectional LSTM (BiLSTM) layers, were explored, among which the two best ones are presented. The DL architectures are summarized in Table 3, with the number of neurons, units, or dropout (D) percentages indicated in parentheses. For example, the architecture of BiLSTM-Net EL model can be seen in Fig. 5. The output layer is a seq2seq regression layer utilizing the loss function outlined in Equation (16).

$$L_{regress} = \frac{1}{2N_j}\sum_{i=1}^{N_j}\left(T_j(t_i) - HI_j(t_i)\right)^2 ; \quad j \in M \quad (16)$$

5. Results and discussions

To thoroughly evaluate various combinations and confirm the effectiveness, validity, and stability of the proposed framework, all potential dataset folds were examined. In each fold, one specimen served as the test set, while the remaining composite specimens were used for training and validation. For both the ReMAP and NASA datasets, which offered four and three-two alternatives for validation, respectively, the process was repeated with 10 iterations using distinct random seed numbers. This approach essentially mirrors the LOOCV, employing a holdout validation strategy within each fold to assess the model's generalizability. The runs were conducted on a high-performance computing cluster, utilizing 20 processors on a single node. In this section, we initially present the results of the HT-SSCNN up to the ensemble stage, referred to as the Base learner models, before delving into the outcomes of the Ensemble learner models.

5.1. Base learner models

The comprehensive results of HT-SSCNN across various subsets and excitation frequencies for T-single stiffener CFRP panels (ReMAP dataset) are presented in Fig. 6. The results display fitness scores (based on Eq. (4)) across different replications and their mean values (indicated on the right y-axis) over ten replications (shown on the x-axis) for each subset (validation index-based combinations displayed on the left y-axis), with error bars representing the standard deviation. The impact of the choice of the validation unit varies depending on the specific test unit and frequency. For instance, in fold 5, where the test unit is specimen 5, unit 1 does not serve as a suitable validation case for GW signals at the excitation frequency of 50 kHz. This is either because it leads to the exclusion of unit 1 information during the training phase or it proves to be an unsuitable validation specimen for terminating the training process. Referring to Fig. 6, it becomes evident that frequencies of 150 kHz and 200 kHz led to more consistent high fitness scores, whereas 100

Table 5
Fitness values for base learner models averaged over the 30 repetitions for NASA dataset (Layup 1). Supplementary tables for other layups in the NASA dataset can be found in the Supplementary Materials.

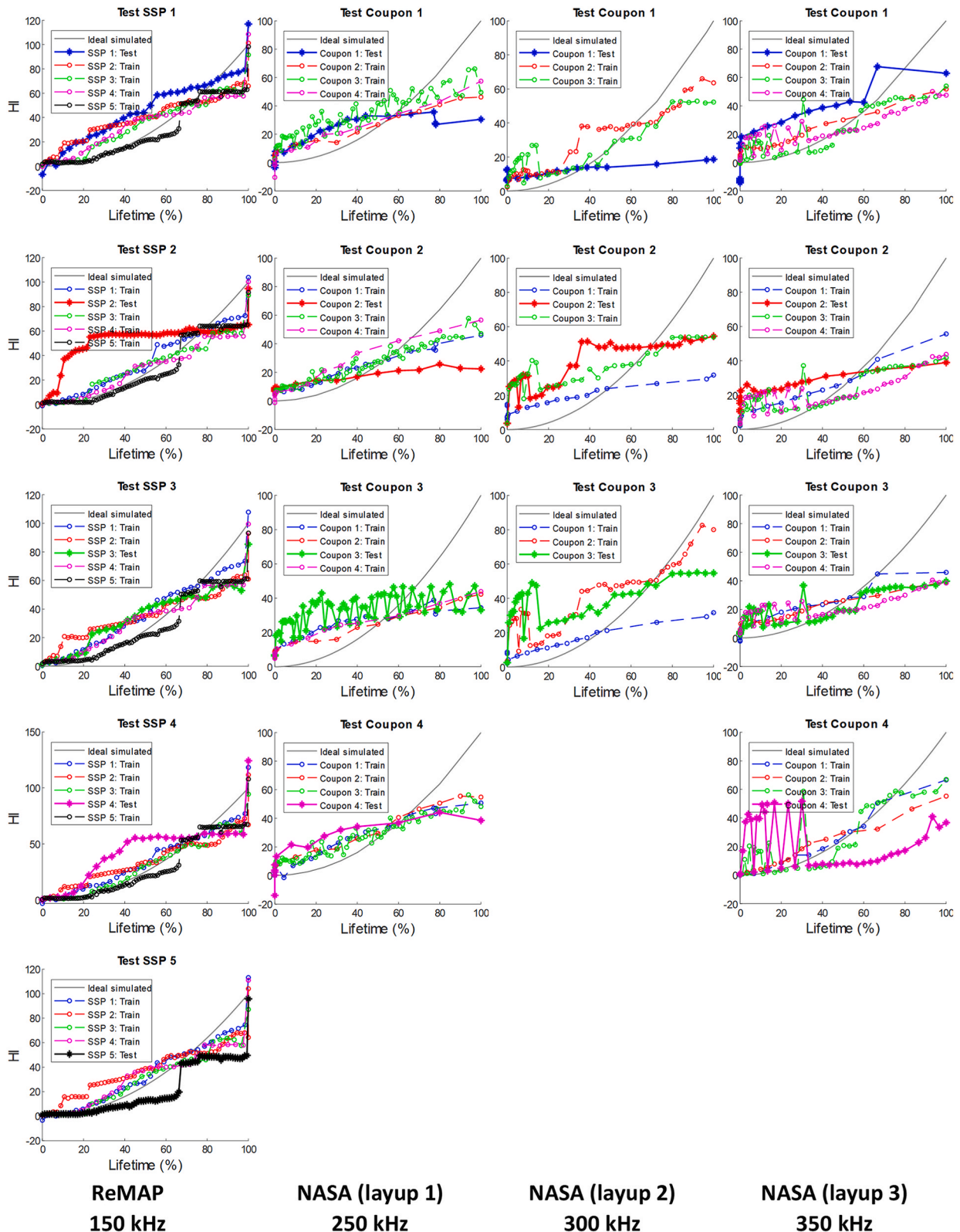| Frequency	Fold 1		Fold 2		Fold 3		Fold 4	
	F-All	F-Test	F-All	F-Test	F-All	F-Test	F-All	F-Test
150 kHz	1.86 ($\pm$0.36)	1.16 ($\pm$0.47)	2.18 ($\pm$0.42)	2.08 ($\pm$0.47)	1.67 ($\pm$0.29)	1.21 ($\pm$0.38)	2.32 ($\pm$0.12)	2.18 ($\pm$0.16)
200 kHz	1.99 ($\pm$0.4)	1.4 ($\pm$0.49)	2.49 ($\pm$0.32)	2.21 ($\pm$0.36)	1.52 ($\pm$0.32)	0.86 ($\pm$0.51)	2.4 ($\pm$0.22)	2.12 ($\pm$0.45)
250 kHz	2.41 ($\pm$0.24)	2.03 ($\pm$0.31)	2.3 ($\pm$0.35)	1.87 ($\pm$0.32)	2.22 ($\pm$0.16)	1.92 ($\pm$0.22)	**2.54 ($\pm$0.26)**	**2.32 ($\pm$0.32)**
300 kHz	2.47 ($\pm$0.2)	2.11 ($\pm$0.24)	2.56 ($\pm$0.15)	2.14 ($\pm$0.27)	1.98 ($\pm$0.22)	1.61 ($\pm$0.28)	2.52 ($\pm$0.2)	2.13 ($\pm$0.33)
350 kHz	2.12 ($\pm$0.35)	1.71 ($\pm$0.39)	1.9 ($\pm$0.39)	1.43 ($\pm$0.52)	1.84 ($\pm$0.25)	1.45 ($\pm$0.25)	2.11 ($\pm$0.41)	1.57 ($\pm$0.75)
400 kHz	2.02 ($\pm$0.29)	1.52 ($\pm$0.58)	1.85 ($\pm$0.46)	1.24 ($\pm$0.51)	1.73 ($\pm$0.28)	1.18 ($\pm$0.42)	2.12 ($\pm$0.34)	1.91 ($\pm$0.53)
450 kHz	2.35 ($\pm$0.38)	2.05 ($\pm$0.44)	2 ($\pm$0.42)	1.59 ($\pm$0.55)	1.71 ($\pm$0.36)	1.25 ($\pm$0.47)	2.26 ($\pm$0.35)	1.87 ($\pm$0.64)
All	2.23 ($\pm$0.35)	1.86 ($\pm$0.48)	2.36 ($\pm$0.3)	2.06 ($\pm$0.28)	2.07 ($\pm$0.4)	1.65 ($\pm$0.43)	2.37 ($\pm$0.32)	2.28 ($\pm$0.48)

Fig. 7. HIs obtained by the proposed HT-SSCNN methodology (with SAE) for different datasets, given single frequency input. The actuating GW frequencies were selected based on the best fitness scores (Tables 4–5 and those in the Supplementary Materials).

kHz and 250 kHz exhibit unstable performance. The Supplementary Materials also include analogous illustrations for the NASA dataset, highlighting the influence of validation unit selection and GW excitation frequency. Tables 4–5 present averaged fitness values (±standard deviation) across all repetitions, irrespective of the validation unit chosen,

for various folds. F-All corresponds to the fitness scores obtained from Eq. (4) considering all units, while F-Test relates to the fitness scores obtained from Eq. (7) focusing solely on the test unit. Tables (similar to Tables 4 and 5), including results for Layups 2 and 3 of the NASA dataset, are provided in the Supplementary Materials.

Table 6
Fitness values for ensemble learner models averaged over the 5 Folds for ReMAP dataset.

| Freq. (kHz) | SAE | | WAE | | | | | | *Deep learning* | | | |
| | | | MSE | | RMSE | | Fitness | | FC | | BiLSTM | |
	F-All	F-Test	F-All	F-Test	F-All	F-Test	F-All	F-Test	F-All	F-Test	F-All	F-Test
50	2.58	2.32	2.58	2.33	2.58	2.32	2.58	2.32	2.60	2.29	2.00	1.64
	(±0.24)	(±0.60)	(±0.21)	(±0.56)	(±0.23)	(±0.59)	(±0.24)	(±0.60)	(±0.16)	(±0.44)	(±0.22)	(±0.16)
100	2.45	2.13	2.25	2.04	2.34	2.08	2.46	2.12	2.62	2.33	2.31	1.97
	(±0.26)	(±0.52)	(±0.44)	(±0.53)	(±0.37)	(±0.51)	(±0.26)	(±0.52)	(±0.26)	(±0.47)	(±0.34)	(±0.52)
125	2.69	2.52	2.66	2.50	2.68	2.51	2.69	2.52	2.68	2.50	2.24	1.75
	(±0.12)	(±0.26)	(±0.12)	(±0.26)	(±0.12)	(±0.27)	(±0.12)	(±0.27)	(±0.13)	(±0.22)	(±0.22)	(±0.20)
150	**2.73 (**	**2.60 (**	2.70	2.58	2.72	2.59	**2.73 (**	**2.60 (**	2.71	2.45	2.43	2.00
	±0.12)	**±0.11)**	(±0.11)	(±0.09)	(±0.12)	(±0.1)	**±0.12)**	**±0.11)**	(±0.14)	(±0.35)	(±0.39)	(±0.36)
200	2.60	2.51	2.57	2.47	2.57	2.48	2.60	2.51	2.60	2.49	2.49	2.18
	(±0.14)	(±0.38)	(±0.16)	(±0.41)	(±0.17)	(±0.41)	(±0.14)	(±0.38)	(±0.15)	(±0.37)	(±0.24)	(±0.52)
250	2.54	2.25	2.49	2.27	2.52	2.26	2.54	2.25	2.49	2.10	2.59	2.26
	(±0.14)	(±0.35)	(±0.12)	(±0.31)	(±0.13)	(±0.34)	(±0.14)	(±0.32)	(±0.18)	(±0.32)	(±0.12)	(±0.46)
Fused (all)	**2.77 (**	**2.66 (**	2.71	2.57	2.75	2.64	**2.78 (**	**2.67 (**	**2.76 (**	**2.62 (**	2.38	2.010
	±0.15)	**±0.22)**	(±0.23)	(±0.40)	(±0.18)	(±0.26)	**±0.15)**	**±0.20)**	**±0.14)**	**±0.22)**	(±0.37)	(±0.43)

Table 7
Fitness values for ensemble learner models averaged over the 4 Folds for NASA dataset (Layup 1). Supplementary tables for other layups in the NASA dataset can be found in the Supplementary Materials.

| Freq. (kHz) | SAE | | WAE | | | | | | *Deep learning* | | | |
| | | | MSE | | RMSE | | Fitness | | FC | | BiLSTM | |
	F-All	F-Test	F-All	F-Test	F-All	F-Test	F-All	F-Test	F-All	F-Test	F-All	F-Test
150	2.14	1.71	2.05	1.73	2.09	1.73	2.15	1.72	1.78	1.43	2.23	1.87
	(±0.37)	(±0.59)	(±0.28)	(±0.51)	(±0.31)	(±0.53)	(±0.37)	(±0.59)	(±0.54)	(±0.92)	(±0.41)	(±0.23)
200	2.24	1.80	2.22	1.86	2.23	1.85	2.25	1.79	2.12	1.54	2.14	1.77
	(±0.52)	(±0.71)	(±0.47)	(±0.71)	(±0.48)	(±0.70)	(±0.52)	(±0.70)	(±0.53)	(±0.90)	(±0.56)	(±0.91)
250	**2.52 (**	**2.12 (**	**2.50 (**	**2.12 (**	**2.50 (**	**2.12 (**	**2.52 (**	**2.13 (**	2.45	2.10	2.33	1.89
	±0.23)	**±0.26)**	**±0.22)**	**±0.26)**	**±0.23)**	**±0.25)**	**±0.21)**	**±0.23)**	(±0.24)	(±0.39)	(±0.21)	(±0.33)
300	2.47	2.06	2.41	2.06	2.44	2.07	2.48	2.08	2.36	1.96	2.34	1.84
	(±0.36)	(±0.32)	(±0.35)	(±0.36)	(±0.35)	(±0.33)	(±0.36)	(±0.32)	(±0.37)	(±0.33)	(±0.44)	(±0.52)
350	2.19	1.82	2.16	1.85	2.18	1.85	2.20	1.82	1.93	1.26	1.78	1.24
	(±0.41)	(±0.52)	(±0.38)	(±0.51)	(±0.39)	(±0.49)	(±0.40)	(±0.52)	(±0.44)	(±0.51)	(±0.42)	(±0.41)
400	2.11	1.70	2.06	1.73	2.09	1.73	2.11	1.69	1.89	1.26	1.73	1.18
	(±0.38)	(±0.52)	(±0.32)	(±0.46)	(±0.35)	(±0.48)	(±0.36)	(±0.51)	(±0.41)	(±0.75)	(±0.18)	(±0.31)
450	2.37	1.95	2.33	1.99	2.37	2.00	2.36	1.93	2.01	1.58	2.04	1.73
	(±0.45)	(±0.62)	(±0.45)	(±0.68)	(±0.45)	(±0.68)	(±0.45)	(±0.57)	(±0.56)	(±0.63)	(±0.46)	(±0.74)
Fused (all)	2.39	2.02	2.35	2.01	2.36	2.03	**2.41 (**	**2.03 (**	2.29	1.92	**2.68 (**	**2.31 (**
	(±0.44)	(±0.64)	(±0.39)	(±0.57)	(±0.41)	(±0.61)	**±0.43)**	**±0.63)**	(±0.50)	(±0.60)	**±0.20)**	**±0.36)**

The highest-scoring frequency is 150 kHz for the ReMAP dataset, with an average of 2.53 for all units and 2.26 for test units across all folds. In the NASA datasets, the top-performing frequencies are 250 kHz, 300 kHz, and 350 kHz for Layups 1, 2, and 3, respectively. The average scores across all folds are 2.37, 2.19, and 2.32 when considering all units, and 2.04, 1.97, and 2.04 when considering test units, as highlighted in green in the tables. Layup 2 yields lower scores because it has limited training data from only one unit.

The results vary across different folds, with the folds showing the highest fitness scores highlighted in bold in the tables. The HIs generated from HT-SSCNN for the optimal frequencies in different datasets are displayed in Fig. 7. These HIs represent the averages across all repetitions, irrespective of the validation unit selection (SAE method). However, showcasing the HIs from a single repetition could yield more promising results.

For T-single stiffener CFRP panels, the results are highly promising despite the presence of uncertainties such as broken sensors, impacts, and disbond. In the NASA dataset, Layup 2 is hindered by a lack of training data, while the HIs for Layup 1 (except for unit 2) and Layup 3 (except for units 1 and 4) are acceptable.

5.2. Ensemble learner models

Ensemble learners are employed to blend diverse predictions and

enhance the effectiveness of HIs compared to the base learner models. Tables 6–7 present the ensemble models' fitness scores averaged across all folds for various ensemble techniques. EL models can be applied to the HIs generated by HT-SSCNN using a single excitation GW frequency or a fusion of all frequencies (labeled '*Fused (all)*' in Tables 6–7). These tables includes results for ReMAP dataset and Layup 1 of the NASA dataset, while results for Layups 2 and 3 of the NASA dataset are provided in the Supplementary Materials.

For the ReMAP dataset, the WAE-Fitness model using the fusion of all GW excitation frequencies achieved the best fitness scores whether considering all or only test units, with F-All of 2.78 and F-Test of 2.67. Additionally, SAE and the FC network (FC-EL) models using the fusion of frequencies also resulted in high scores. Aside from the fusion of all frequencies, the 150 kHz frequency, when employed with SAE and WAE-Fitness, yielded high fitness scores (exceeding 2.6, meeting 87% of evaluation criteria).

In the NASA dataset, for Layup 1, the BiLSTM network (BiLSTM-EL) using the fusion of all frequencies obtained the highest fitness scores (F-All of 2.68 and F-Test of 2.31). However, it is important to note that this HI construction model for Layup 1 is not historical-independent as it uses prior information in the EL step. WAE-Fitness with GW signals generated from the 250 kHz frequency achieved better fitness scores in second place. The WAE-Fitness using the fused frequencies for this layup produces scores that are not very low (F-All of 2.41 and F-Test of 2.03).
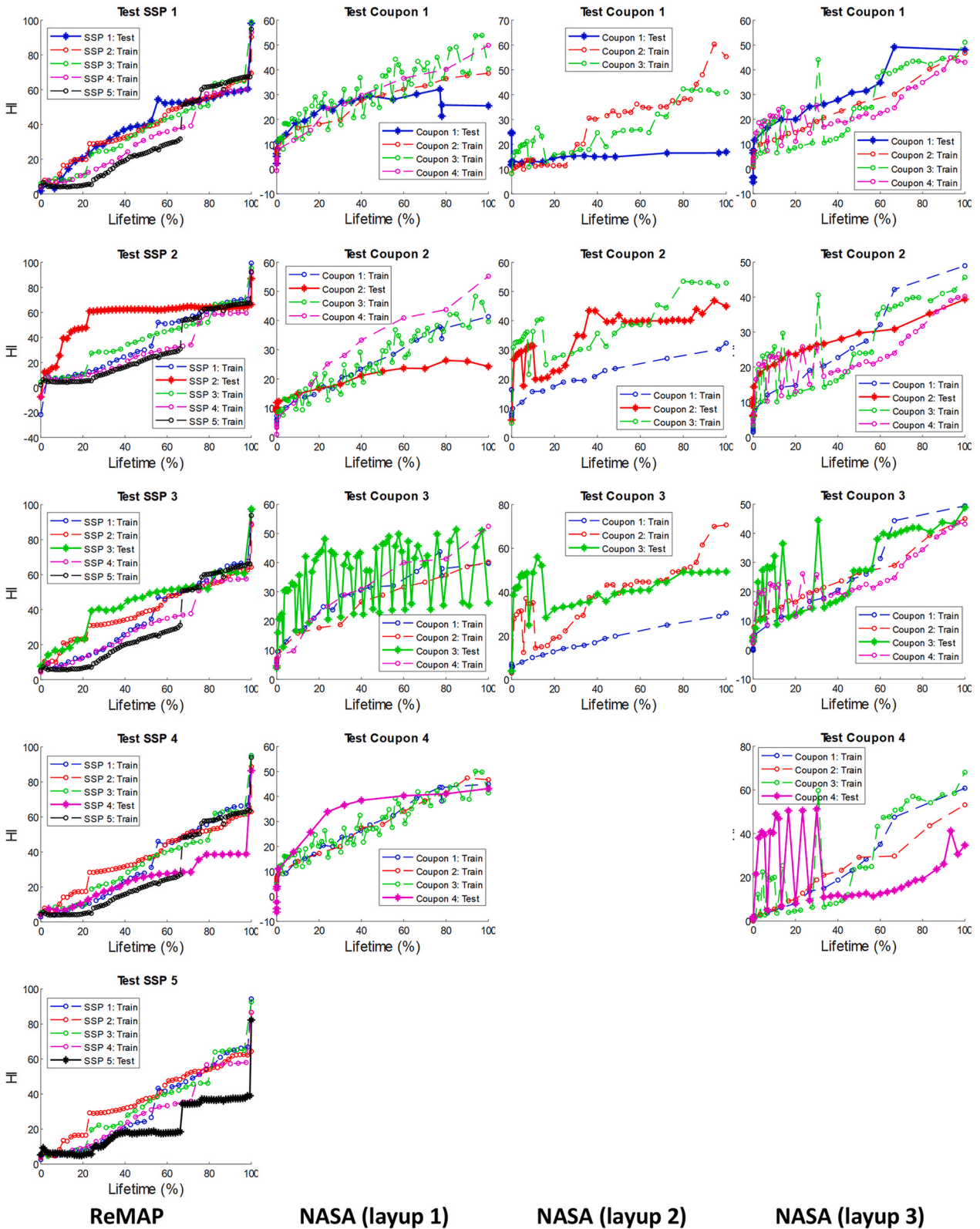
Fig. 8. HIs obtained by the proposed methodology with WAE-Fitness for different datasets, given all frequency inputs.

For Layup 2, EL models on the GW excitation frequency of 300 kHz provided higher scores than others, but the fitness scores are not yet highly proper (more than 90% performance, i.e., above 2.7) as this dataset, as previously mentioned, suffers from the limited data needed to train the models. In Layup 3, WAE-Fitness on the fusion of all frequencies achieved the highest fitness scores (F-All of 2.42 and F-Test of

2.23).

On average, WAE-Fitness using the fusion of all GW excitation frequencies resulted in higher fitness scores, and the related constructed HIs using this model (i.e., the end-to-end model made of HT–SSCNN–WAE$_{Fitness}$) can be seen in Fig. 8. Considering the scores reported in Table 6 and the generated HIs in Fig. 8 for T-single CFRP

panels under C–C fatigue loading (ReMAP dataset), the performance is highly satisfactory, achieving 93% (2.78/3.00) given all units and 89% (2.67/3.00) given test units. The HIs exhibit monotonic, prognostable, and trendable behavior, making them suitable for predicting the RUL of composite structures.

For dogbone CFRP specimens under T-T fatigue loading (NASA dataset), layups with four units (Layup 1 and Layup 3), which have 2 units for training and 1 unit for validation, yield better results compared to Layup 2, which has only one unit for training and another for validation. This can be observed in Fig. 8, particularly for coupon 4 of Layup 1 and coupons 1, 2, and 3 of Layup 3.

HIs constructed by BiLSTM-EL upon HT-SSCNN can be seen in the Supplementary Materials. These HIs interestingly exhibit multiple incremental steps over the fatigue life in the ReMAP dataset and Layup 1 of the NASA dataset. These steps may signify distinct damage states, providing valuable insights for subsequent prognostic models, especially state-based ones, for RUL prediction. However, establishing a meaningful link between these steps and physical damage states requires further research and experimentation. The HIs constructed by FC-EL after the HT-SSCNN model can also be found in the Supplementary Materials.

In general, higher fitness scores for HIs could have been attained for the ReMAP datasets compared to the NASA dataset. We attribute this to several factors, including the larger number of training specimens (only one unit more) and a greater number of time steps for GW inspections, which provide more training data for the DL models. This difference in performance seems to be less related to the structural type, loading conditions, layups, or types of damage, as the T-single stiffener panels are inherently more complex in various aspects. Moreover, the ReMAP structures were monitored using a more intertwined GW network with 56 paths, while the NASA structures used 36 paths. This richer, intertwined GW network offers more information to leverage for model training. In essence, the model's performance is less affected by issues encountered during the monitoring process, such as the presence of broken or debonded PZT sensors, which occurred during ReMAP experiments. It should be noted that data augmentation techniques may address the limited number of units available to train the model, such as the NASA dataset, and can improve performance, where data size for machine learning needs to be optimized after sensitivity analysis of the size [53].

6. Conclusions

In this study, we introduced an innovative approach to construct comprehensive health indicators (HIs) for composite structures, addressing the challenges posed by the stochastic nature of damage accumulation during operational conditions and the need for HIs independent of historical data. Leveraging the power of AI, we developed the Hilbert transform-convolutional neural network (HT-SSCNN) within a semi-supervised learning paradigm. The approach exclusively utilizes current guided wave (GW) data, eliminating the reliance on historical information. It flexibly accommodates different GW sensor numbers, networks, and setups. The results demonstrate the effectiveness and validity of the approach. To assess various combinations and ensure robustness, rigorous evaluations were conducted, considering different datasets under various conditions.

Our findings indicate that certain frequencies, such as 150 kHz for the ReMAP dataset and 250 kHz for NASA Layup 1, consistently outperformed others, resulting in more stable and reliable HIs. The use of ensemble learning techniques, specifically WAE-Fitness, led to significant improvements in HIs' performance. For the ReMAP dataset, the WAE-Fitness model, fusing all GW excitation frequencies, yielded the best fitness scores, with 93% accuracy considering all units and 89% given test units. While ReMAP experiments are more complex than the NASA ones, HIs with higher performance could be extracted from the ReMAP dataset than the NASA dataset, which can be attributed to the availability of one more training specimen and a greater number of time steps for GW inspections. Additionally, ReMAP structures were monitored using a more intertwined GW sensory network, which provided a wealth of data for model training. The produced HIs exhibit desirable properties for RUL prediction. They are monotonic, prognostable, and exhibit correlated trends, which are essential characteristics for accurate predictions in PHM. The incremental steps observed in the HIs may potentially correspond to distinct damage states, which can be used to inform future state-based RUL prediction models. In conclusion, our approach offers a promising solution to the challenging task of constructing reliable and historical-independent HIs for composite structures. By combining AI with signal processing algorithms, we achieved a high level of performance and demonstrated the applicability of our method across different datasets. This research opens up new possibilities for enhanced health monitoring and predictive maintenance of composite structures in various applications, including aerospace engineering.

Funding

CRediT authorship contribution statement

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

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compositesb.2024.111328.

References

[1] Peng T, Liu Y, Saxena A, Goebel K. In-situ fatigue life prognosis for composite laminates based on stiffness degradation. Compos Struct Nov. 2015;132:155–65. https://doi.org/10.1016/J.COMPSTRUCT.2015.05.006.

[2] Saeedifar M, Zarouchas D. Damage characterization of laminated composites using acoustic emission: a review. Compos B Eng Aug. 2020;195:108039. https://doi.org/10.1016/J.COMPOSITESB.2020.108039.

[3] Giurgiutiu V. Structural health monitoring of aerospace composites. Elsevier; 2015. https://doi.org/10.1016/B978-0-12-409605-9.00012-X.

[4] Shabani P, Taheri-Behrooz F, Samareh-Mousavi SS, Shokrieh MM. Very high cycle and gigacycle fatigue of fiber-reinforced composites: a review on experimental approaches and fatigue damage mechanisms. Prog Mater Sci May 2021;118: 100762. https://doi.org/10.1016/J.PMATSCI.2020.100762.

[5] Ahmed O, Wang X, Tran MV, Ismadi MZ. Advancements in fiber-reinforced polymer composite materials damage detection methods: towards achieving energy-efficient SHM systems. Compos B Eng Oct. 2021;223:109136. https://doi.org/10.1016/J.COMPOSITESB.2021.109136.

[6] Guida M, Marulo F, Belkhelfa FZ, Russo P. A review of the bird impact process and validation of the SPH impact model for aircraft structures. Prog Aero Sci Feb. 2022; 129:100787. https://doi.org/10.1016/J.PAEROSCI.2021.100787.

[7] Farrar CR, Lieven NAJ. Damage prognosis: the future of structural health monitoring. Phil Trans Math Phys Eng Sci Dec. 2006;365(1851):623–32. https://doi.org/10.1098/RSTA.2006.1927.

[8] Lee IY, Doh Roh H, Bin Park Y. Prognostics and health management of composite structures under multiple impacts through electromechanical behavior and a particle filter. Mater Des Nov. 2022;223:111143. https://doi.org/10.1016/J.MATDES.2022.111143.

[9] Coble JB. Merging data Sources to predict remaining useful life – an Automated method to identify prognostic parameters. Doctoral Dissertations; May 2010. Sep. 12, 2023. [Online]. Available: https://trace.tennessee.edu/utk_graddiss/683.

[10] Guo L, Li N, Jia F, Lei Y, Lin J. A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing 2017;240: 98–109. https://doi.org/10.1016/j.neucom.2017.02.045.

[11] Saxena A, et al. Metrics for evaluating performance of prognostic techniques. In: 2008 International Conference on prognostics and health management. PHM 2008; 2008. https://doi.org/10.1109/PHM.2008.4711436.

[12] Lei Y. Intelligent fault diagnosis and remaining useful life prediction of rotating machinery. 2016. https://doi.org/10.1016/c2016-0-00367-4.

[13] Moradi M, Chiachío J, Zarouchas D. Developing health indicators for composite structures based on a two-stage semi-supervised machine learning model using acoustic emission data. In: The 10th ECCOMAS Thematic Conference on Smart structures and materials; 2023. Sep. 15, 2023. [Online]. Available: https://research.tudelft.nl/en/publications/developing-health-indicators-for-composite-structures-based-on-a-.

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

[15] Ihn JB, Chang FK. Pitch-catch active sensing methods in structural health monitoring for aircraft structures. Struct Health Monit Mar. 2008;7(1):5–19. https://doi.org/10.1177/1475921707081979.

[16] Zhang B, Yang D, Hong X, Jin G. Deep emulational semi-supervised knowledge probability imaging method for plate structural health monitoring using guided waves. Eng Comput Oct. 2022;38(5):4151–66. https://doi.org/10.1007/S00366-022-01711-9/TABLES/3.

[17] Tang L, et al. Quantitative identification of damage in composite structures using sparse sensor arrays and multi-domain-feature fusion of guided waves. Measurement Feb. 2023;208:112482. https://doi.org/10.1016/J.MEASUREMENT.2023.112482.

[18] Yue N, Broer A, Briand W, Rébillat M, Loutas T, Zarouchas D. Assessing stiffness degradation of stiffened composite panels in post-buckling compression-compression fatigue using guided waves. Compos Struct Aug. 2022;293:115751. https://doi.org/10.1016/J.COMPSTRUCT.2022.115751.

[19] Su Z, Ye L, Lu Y. Guided Lamb waves for identification of damage in composite structures: a review. J Sound Vib Aug. 2006;295(3–5):753–80. https://doi.org/10.1016/j.jsv.2006.01.020.

[20] De Luca A, Caputo F, Sharif Khodaei Z, Aliabadi MH. Damage characterization of composite plates under low velocity impact using ultrasonic guided waves. Compos B Eng Apr. 2018;138:168–80. https://doi.org/10.1016/J.COMPOSITESB.2017.11.042.

[21] Memmolo V, Monaco E, Boffa ND, Maio L, Ricci F. Guided wave propagation and scattering for structural health monitoring of stiffened composites. Compos Struct Jan. 2018;184:568–80. https://doi.org/10.1016/j.compstruct.2017.09.067.

[22] Gorgin R, Luo Y, Wu Z. Environmental and operational conditions effects on Lamb wave based structural health monitoring systems: a review. Ultrasonics Jul. 2020; 105. https://doi.org/10.1016/j.ultras.2020.106114.

[23] Santoni GB, Yu L, Xu B, Giurgiutiu V. Lamb wave-mode tuning of piezoelectric wafer active sensors for structural health monitoring. J Vib Acoust 2007;129(6): 752–62. https://doi.org/10.1115/1.2748469.

[24] Larrosa C, Lonkar K, Chang FK. In situ damage classification for composite laminates using Gaussian discriminant analysis. Struct Health Monit Mar. 2014;13 (2):190–204. https://doi.org/10.1177/1475921713517288.

[25] Yan G, Lu X, Tang J. Guided wave-based monitoring of evolution of fatigue damage in glass Fiber/Epoxy composites. Appl Sci Apr. 2019;9(7). https://doi.org/10.3390/app9071394.

[26] A. Saxena, K. Goebel, C. C. Larrosa, V. Janapati, S. Roy, and F.-K. Chang, "Accelerated aging experiments for prognostics of damage Growth in composite materials.".

[27] Liu H, Liu S, Liu Z, Mrad N, Milani AS. Data-driven approaches for characterization of delamination damage in composite materials. IEEE Trans Ind Electron Mar. 2021;68(3):2532–42. https://doi.org/10.1109/TIE.2020.2973877.

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

[29] Miorelli R, Fisher C, Kulakovskyi A, Chapuis B, Mesnil O, D'Almeida O. Defect sizing in guided wave imaging structural health monitoring using convolutional neural networks. NDT E Int 2021;122(Sep). https://doi.org/10.1016/j.ndteint.2021.102480.

[30] Tabatabaeian A, Jerkovic B, Harrison P, Marchiori E, Fotouhi M. Barely visible impact damage detection in composite structures using deep learning networks with varying complexities. Compos B Eng Sep. 2023;264:110907. https://doi.org/10.1016/J.COMPOSITESB.2023.110907.

[31] Moradi M, Broer A, Chiachío J, Benedictus R, Loutas TH, Zarouchas D. Intelligent health indicator construction for prognostics of composite structures utilizing a semi-supervised deep neural network and SHM data. Eng Appl Artif Intell Jan. 2023;117:105502. https://doi.org/10.1016/J.ENGAPPAI.2022.105502.

[32] Zarouchas D, Broer A, Galanopoulos G, Briand W, Benedictus R, Loutas T. Compression Compression fatigue tests on single stiffener aerospace structures. 2021. https://doi.org/10.34894/QNURER.

[33] Saxena A, Goebel KF, Larrosa CC, Janapati V, Roy S, Chang F-K. Accelerated aging experiments for prognostics of damage Growth in composite materials. 2011.

[34] Coble J, Hines JW. Identifying optimal prognostic parameters from data: a Genetic algorithms approach. Annual Conference of the PHM Society 2009;1(1). Sep. 16, 2023. [Online]. Available: http://www.papers.phmsociety.org/index.php/phmconf/article/view/1404.

[35] Eleftheroglou N, Zarouchas D, Loutas T, Alderliesten R, Benedictus R. Structural health monitoring data fusion for in-situ life prognosis of composite structures. Reliab Eng Syst Saf Oct. 2018;178:40–54. https://doi.org/10.1016/J.RESS.2018.04.031.

[36] Chiachio M, Chiachío J, Saxena A, Goebel K. Documentation for the fatigue dataset in composites. Sep. 2013. Oct. 29, 2023. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Documentation+for+the+fatigue+dataset+in+composites&btnG=.

[37] Saxena A, Goebel K, Larrosa CC, Chang FK. CFRP Composites dataset. 2008. Oct. 29, 2023. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Cfrp+composites+dataset%2C+nasa+ames+prognostics+data+repository&btnG=.

[38] Rose JL. Ultrasonic guided waves in Solid Media. Ultrasonic Guided Waves in Solid Media Jan. 2014;9781107048959:1–512. https://doi.org/10.1017/CBO9781107273610.

[39] Oppenheim AV. Discrete-time signal processing. Oct. 27, 2023. [Online]. Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=Oppenheim%2C+A.+V.+%281999%29.Discrete-time+signal+processing.+Pearson+Education+India.+&btnG=; 1999.

[40] Staszewski WJ, Lee BC, Mallet L, Scarpa F. Structural health monitoring using scanning laser vibrometry: I. Lamb wave sensing. Smart Mater Struct Feb. 2004;13 (2):251. https://doi.org/10.1088/0964-1726/13/2/002.

[41] Zhao X, Royer RL, Owens SE, Rose JL. Ultrasonic Lamb wave tomography in structural health monitoring. Smart Mater Struct Aug. 2011;20(10):105002. https://doi.org/10.1088/0964-1726/20/10/105002.

[42] Moradi M, Broer A, Chiachío J, Benedictus R, Zarouchas D. Intelligent health indicators based on semi-supervised learning utilizing acoustic emission data. Lecture Notes in Civil Engineering 2023;270:419–28. https://doi.org/10.1007/978-3-031-07322-9_43. LNCE.

[43] Moradi M, Komninos P, Benedictus R, Zarouchas D. Interpretable neural network with limited weights for constructing simple and explainable HI using SHM data. In: Annual Conference of the. PHM Society; 2022. https://doi.org/10.36001/phmconf.2022.v14i1.3185.

[44] Maas A, Hannun A. Rectifier nonlinearities improve neural network acoustic models. A. N.-Proc. icml, and undefined 2013. In: *robotics*.stanford.eduAL Maas, AY Hannun, AY NgProc. icml, 2013•robotics.stanford.edu; 2013. Jan. 19, 2024. [Online]. Available: http://robotics.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf.

[45] Kingma DP, Ba JL. Adam: a method for stochastic optimization," *3rd International Conference on learning Representations*. ICLR 2015 - Conference Track Proceedings; 2015.

[46] Breiman L. Bagging predictors. Mach Learn 1996;24(2):123–40. https://doi.org/10.1007/BF00058655/METRICS.

[47] Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors) Apr. 2000;28(2): 337–407. https://doi.org/10.1214/AOS/1016218223.

[48] Wolpert DH. Stacked generalization. Neural Network Jan. 1992;5(2):241–59. https://doi.org/10.1016/S0893-6080(05)80023-1.

[49] Haykin S. Neural networks: a comprehensive foundation. Prentice Hall PTR; 1998. https://doi.org/10.1016/0967-0661(95)90080-2.

[50] Hashem S. Optimal linear combinations of neural networks. Neural Network Jun. 1997;10(4):599–614. https://doi.org/10.1016/S0893-6080(96)00098-6.

[51] Hastie T, Friedman J, Tibshirani R. The Elements of Statistical Learning 2001;45 (3). https://doi.org/10.1007/978-0-387-21606-5. in Springer Series in Statistics, vol. 45. New York, NY: Springer New York.

[52] Matsumoto M, Nishimura T. Mersenne twister. ACM Trans Model Comput Simulat Jan. 1998;8(1):3–30. https://doi.org/10.1145/272991.272995.

[53] Dabetwar S, Ekwaro-Osire S, Dias JP. Fatigue damage diagnostics of composites using data fusion and data augmentation with deep neural networks. J Nondestruct Eval Diagn Prognos Eng Syst May 2022;5(2). https://doi.org/10.1115/1.4051947/1115173.