Health indicator modeling leveraging time-independent and time-dependent subtasks with adaptive standardization and physics-based Bayesian optimization for aeronautical structures

Moradi, Morteza; Chiachío, Juan; Zarouchas, Dimitrios

DOI
10.1016/j.engappai.2025.112767

Publication date
2026

Document Version
Final published version

Published in
Engineering Applications of Artificial Intelligence

Citation (APA)
Moradi, M., Chiachío, J., & Zarouchas, D. (2026). Health indicator modeling leveraging time-independent and time-dependent subtasks with adaptive standardization and physics-based Bayesian optimization for aeronautical structures. *Engineering Applications of Artificial Intelligence*, *163*, Article 112767. https://doi.org/10.1016/j.engappai.2025.112767

Engineering Applications of Artificial Intelligence 163 (2026) 112767

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Health indicator modeling leveraging time-independent and time-dependent subtasks with adaptive standardization and physics-based Bayesian optimization for aeronautical structures

Morteza Moradi [a,*], Juan Chiachío [b], 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

[b] Dept. Structural Mechanics & Hydraulics Engineering, Andalusian Research Institute in Data Science and Computational Intelligence (DaSCI), University of Granada, Granada, 18001, Spain

ARTICLE INFO

Keywords:
Prognostics and health management
Structural health monitoring
Intelligent health indicator
Semi-supervised deep neural network
Composite structures

ABSTRACT

Monitoring the structural integrity of aeronautical structures is critical for safety, reducing maintenance costs, and enabling predictive maintenance. However, raw structural health monitoring (SHM) data are often noisy, high-dimensional, and difficult to interpret. To enable condition-based maintenance, it is essential to extract health indicators (HIs)—quantitative representations of structural degradation that evolve consistently over time. Accurately extracting HIs for composite structures is particularly challenging due to complex material behavior and multiple damage sources. While deep learning models offer potential, their application is limited by the lack of run-to-failure data and ground-truth HI labels. To address these challenges, this study proposes a novel approach that divides HI modeling into two tasks: time-independent (spatial) and time-dependent (temporal). This separation allows more effective data utilization, especially in the time-independent case. A semi-supervised spatial model is first developed and fine-tuned using a Bayesian algorithm with a coupled physics-based loss function that integrates both prognostic criteria and simulated labels—explicitly through the former and implicitly through the latter—embedding degradation physics into training. The study also introduces a new adaptive standardization technique for fatigue-based SHM and systematically evaluates principal component analysis (PCA)-based methods for dimensionality reduction prior to spatial and temporal modeling, simplifying subsequent network architectures. In the final stage, following time-based resampling, a semi-supervised temporal model captures HI evolution, with ensemble learning enhancing robustness. Validation on single-stiffener composite panels under fatigue loading, monitored via acoustic emission sensors, confirms the framework's generalizability and performance—achieving up to 90% (±2%) accuracy in prognostic metrics.

1. Introduction

The use of composite materials as primary structural components has become widespread in aerospace applications over the past decade, driven by their high strength-to-weight ratio, corrosion resistance, and fatigue performance (Beaumont, 2020). However, throughout their service life, these structures endure complex loading scenarios and harsh environmental conditions, which complicate the assessment of their mechanical integrity and the prediction of their remaining useful life (RUL) (Ameri et al., 2020), (Peng et al., 2015). The situation becomes even more challenging when unforeseen events such as impacts occur, necessitating conservative design strategies like safety factors to prevent catastrophic failures (Suresh Kumar et al., 2019). While effective at mitigating risk, safety factors often diminish the full potential of composite materials in terms of weight savings and performance. To avoid this trade-off and fully exploit the advantages of composites, more accurate predictive models of structural behavior are required. Because physics-based models struggle to capture long-term composite behavior under operational conditions (Peng et al., 2015), (Wu and Yao, 2010), data-driven approaches have gained prominence (Nuhic et al., 2013), (Ferreira and Gonçalves, 2022).

In this context, structural health monitoring (SHM) plays a pivotal role by enabling continuous in-service data acquisition from operating structures, allowing the extraction of health indicators (HIs)—

* Corresponding author.
E-mail address: M.Moradi-1@tudelft.nl (M. Moradi).

https://doi.org/10.1016/j.engappai.2025.112767
Received 24 January 2025; Received in revised form 22 September 2025; Accepted 10 October 2025
Available online 31 October 2025

quantitative representations of structural degradation over time (Moradi et al., 2022a), (Guo et al., 2017). HIs serve as a critical interface between diagnostics (damage identification) and prognostics (damage evolution and failure prediction), forming the foundation of condition-based maintenance (CBM) strategies (Nuhic et al., 2013), (Ferreira and Gonçalves, 2022), (El-Thalji and Jantunen, 2015). In aerospace systems, where reliability is paramount, accurate and comprehensive HIs are essential for early warning and timely maintenance decisions before irreversible damage occurs (Beaumont, 2020), (Moradi, 2024), (Moradi et al., 2024a), (Yu and Li, 2022), (Fu and Avdelidis, 2023). However, constructing effective HIs for complex systems—especially those made of composite materials—remains challenging. A truly comprehensive HI must account for all relevant degradation mechanisms, many of which are not directly observable due to the hidden nature of damage and limitations of sensor technology (Beaumont, 2020), (Moradi, 2024), (Moradi et al., 2024b). The effects of multiple damage modes may be captured in sensed signals, but decoupling and interpreting these signals is often confounded by numerous known and unknown variables, including sensor type, placement, and the SHM technique employed (Kralovec and Schagerl, 2020), (Farrar and Worden, 2007). Furthermore, SHM data from composite fatigue tests are often expensive to obtain and include many unlabeled (pre-failure) measurements, limiting the applicability of purely supervised learning approaches (Yuan et al., 2020), (Moradi et al., 2023a). As such, asserting the availability of true or complete HI values is rarely justifiable, particularly in the context of high-performance aerospace structures (Beaumont, 2020), (Moradi, 2024), (Saxena et al., 2021). These constraints motivate the need for novel modeling strategies that can learn meaningful HIs from incomplete labels and evolving sensor data.

To address these challenges, this study introduces a two-stage HI modeling framework tailored for fatigue-damaged aerospace composites. First, extracted features from sensor data (here, acoustic emission data) are adaptively standardized using statistics accumulated up to each time point, accounting for unit-specific baselines and non-stationarity. Then, principal component analysis (PCA) reduces dimensionality, yielding a compact set of informative features. A time-independent neural network sub-model regresses these PCA components to a first-level HI. A time-dependent sub-model then refines this HI by explicitly modeling its temporal evolution. By decoupling data-rich, time-independent learning from data-scarce, time-dependent learning, the framework maximizes data utilization from the spatial domain and enforces temporal consistency.

Key contributions and innovations of this study include:

- Two-stage modeling: We separate HI construction into time-independent and time-dependent subtasks, enabling the use of a larger pool of training samples before refining temporal dynamics.
- Semi-supervised spatial learning: We design a spatial model trained in an inductive semi-supervised manner, leveraging Bayesian optimization with a dual-component physics-based loss function. This loss combines (i) an explicit term enforcing prognostic criteria grounded in degradation physics and (ii) an implicit term minimizing error against simulated HI labels derived from those criteria—enabling robust HI learning when actual failure labels are scarce.
- Adaptive standardization: We introduce a novel adaptive standardization technique tailored for fatigue-based SHM problems. To the best of our knowledge, this is the first study to formalize and implement this approach, particularly in the context of prognostic and HI modeling.
- Comprehensive evaluation of PCA-based methods: We conduct a comprehensive comparison of PCA variants as a fast, unsupervised dimensionality reduction strategy, yielding simpler downstream models with fewer learnable parameters.

Together, these contributions form a robust SHM framework for reliable HI construction in fatigue-loaded composite structures.

The remainder of this paper is structured as follows: Section 2 reviews the relevant literature. Section 3 describes the experimental setup and feature extraction process. Section 4 outlines HI evaluation criteria. In Section 5, we introduce the two-stage semi-supervised, physics-informed fusion model and ensemble strategies. Section 6 presents results and analyses—including comparisons to state-of-the-art methods—and Section 7 concludes the paper with key findings and future research directions.

2. Literature review

It is well established that the mechanical properties of structural materials—including aerospace composites—degrade irreversibly over time, so a robust HI must mirror this behavior. A metric known as monotonicity (Mo) has been proposed, which assesses if the HI sufficiently represents this characteristic (Saxena et al., 2008), (Lei, 2016). Furthermore, as a group of similar *units* (structures) approaches their end-of-life (EoL), their HIs should converge to the same value, representing the failure threshold. However, HIs at the EoL may exhibit fluctuations and fail to converge to a consistent value. This variance can be quantified using a metric called prognosability (Pr). Low variability indicates a reliable failure threshold, which is critical for consistent RUL prediction. Trendability (Tr), in contrast, evaluates how uniformly the HIs of different units evolve over time. A high trendability score implies that damage progression follows similar trends across units, enabling generalization in prognostic modeling (Lei, 2016), (Saidi et al., 2017), (Coble, 2010). While the first two assessment criteria (Mo and Pr) are grounded in physical reality and are generally considered objective, achieving high Tr is particularly challenging in real-world aerospace SHM scenarios due to stochastic damage mechanisms, variable operational loads, and environmental uncertainty. Nevertheless, striving for HIs with high Tr remains a goal to improve the accuracy of RUL predictions. Despite these complexities, constructing HIs that satisfy Mo, Pr, and Tr remains a foundational requirement in prognostics, and they serve as essential criteria for validating any proposed HI modeling approach (Nguyen and Medjaher, 2021), (Baptista et al., 2022). Although indispensable, the joint "fitness" objective Mo + Pr + Tr is intrinsically non-differentiable, and both Pr and Tr can only be computed retrospectively from complete run-to-failure records of multiple units—data that are seldom available during model development. Consequently, most published studies satisfy these criteria only partially or under assumptions that compromise field deployment.

Liu et al. (2009) explored a hybrid sensing approach combining active piezoelectric and acoustic emission (AE) techniques to monitor fatigue degradation in eight-ply unidirectional carbon/epoxy composite beams subjected to three-point bending fatigue. Wavelet-based signal features were extracted from piezoelectric sensors, while AE counts served as a simple, interpretable indicator of accumulated damage. Feature trends were clearly observable over the degradation timeline, particularly in AE counts and low-frequency signal energy of piezoelectric signals. However, feature normalization relied on max–min scaling using full data trajectories—an unrealistic assumption in real-world prognostics. Furthermore, the study did not incorporate physics-based constraints or assess prognostic performance using Mo, Pr, or Tr criteria. Milanoski and Loutas (2021) investigated composite panels with artificially introduced skin–stiffener disbonds. Supported by finite element simulations, fiber Bragg grating (FBG) sensors were used to formulate two strain-based HIs: (i) the relative strain deviation between disbonded and baseline conditions and (ii) the fractional contribution of each sensor to the total cumulative strain (Sbarufatti et al., 2013). Again, prognostic metrics (Mo, Pr, Tr) were not evaluated and no explicit physics-based constraints were incorporated.

Galanopoulos et al. extended FBG-based monitoring by incorporating acoustic emission (AE) information in two sequential studies on single-stiffener composite panels subjected to variable-amplitude compression–compression fatigue. In the first study (Galanopoulos et al.,

2021a), the above-mentioned FBG strain deviations were complemented with two AE-derived HIs—windowed cumulative features such as Hits and rise-time/amplitude ratio (RA) values—each accumulated over 500-cycle intervals. The windowing smooths the intrinsically sparse AE stream, yielding trajectories that are amenable to prognostic interpretation. No physics-based constraints were applied and the prognostic metrics Mo, Pr, and Tr were not quantified. The follow-up investigation (Galanopoulos et al., 2021b) broadened the comparison space to eight candidate indicators. From FBG strain data, the authors extracted four physics-based HIs:

- HI1: Relative strain-based HI, measuring deviation from the baseline (similar to (Milanoski and Loutas, 2021), (Galanopoulos et al., 2021a)).
- HI2: Relative/cumulative strain-based HI, considering the accumulated change across the panel length (similar to (Milanoski and Loutas, 2021), (Sbarufatti et al., 2013), (Galanopoulos et al., 2021a)).
- HI3 and HI4: Monotonicity-weighted fusions of HI1 and HI2, respectively, across all sensors—emphasizing sensor contributions based on their individual trend reliability.

Two additional 'virtual' strain HIs from FBG strain data were constructed via principal component analysis (PCA):

- vHI1: Used PCA to reduce dimensionality (from 10 to 2 components), then calculated Euclidean distance in the reduced space, normalized with a radial basis function to ensure a [1, 0] scale (start to failure). Its main limitation lies in its reliance on known min/max values.
- vHI2 (Q-index): Captured the sum of PCA residual reconstruction errors, sensitive to deviations from the reference healthy state.

From AE data, the same HIs used in (Galanopoulos et al., 2021a)—windowed cumulative Hits and RA—were retained, giving a total of eight indicators. Comparative analysis revealed a consistent trade-off: strain-based HIs (HI1–HI4) exhibited high qualitative monotonicity and trendability but low prognosability, whereas AE-based HIs were more prognosable and less sensitive to load variability, albeit less monotonic. Although the feature design was physics-informed, explicit physical constraints were not embedded in the modeling, and prognostic metrics were only qualitatively discussed, with indications that HI3, HI4, vHI1, and vHI2 performed well in terms of monotonicity.

Yue et al. (2022) estimated stiffness degradation in stiffened composite panels by analyzing guided wave (GW) propagation at three excitation frequencies: 100, 125, and 150 kHz. Among these, the 100 kHz frequency yielded the highest prognostic performance, achieving Mo = 1.00, Pr = 0.60, and Tr = 0.95, resulting in a total score of 2.55 out of 3 (≈85 %). It should be noted that a score of 1.00 is the optimal value for each individual metric, and the maximum fitness score is 3.00. Compared with the reference stiffness obtained from load–displacement measurements, the GW-based HI offered better monotonicity and trendability but poorer prognosability. Prognostic performance proved sensitive to excitation frequency, an operational-monitoring variable that cannot be selected a priori and is likely to vary with structural configuration; a weighted fusion of the three frequency-specific indices reduced the fitness marginally to 2.52. Finally, monotonicity was assessed via a simple slope criterion rather than the more rigorous modified Mann–Kendall (MMK) test, leaving some uncertainty regarding its statistical robustness.

Moradi et al. (2023a) were the first to embed all three metrics implicitly within an intrinsically semi-supervised LSTM loss, yielding a fitness of 2.79. However, their leave-one-out cross-validation (LOOCV) procedure employed test-unit data for early stopping of training—an approach incompatible with real-world scenarios. A genetic programming fusion proposed by Galanopoulos et al. (2023) improved monotonicity (Mo = 0.84) and prognosability (Pr = 0.91) but omitted trendability. More recently, Moradi et al. (2024b) addressed all three criteria and reported high fitness—93% for T-single-stiffener carbon fiber reinforced polymer (CFRP) panels under compression fatigue and 81% for dog-bone CFRP specimens under tension fatigue—using a Hilbert transform semi-supervised convolutional neural network (HT-SSCNN) applied to guided wave data; nonetheless, the approach is constrained by the coarse temporal resolution inherent to guided wave acquisition.

As summarized in Table 1, three key methodological challenges emerge from the literature. (i) HIs are inherently history-dependent (Ahmadivala, 2020), (Valkonen, 2023), (Kamranfar, 2023); however, sequence-to-sequence models struggle due to the limited availability of run-to-failure trajectories, while window-based point models tend to underutilize temporal information. (ii) Unit-to-unit variability—stemming from manufacturing defects, impact locations, and load history—induces feature space drift. Consequently, global normalization introduces bias into the HI, and no previous study has implemented adaptive, per-unit scaling to address this issue. (iii) Windowed AE statistics often result in high-dimensional feature spaces (typically more than 100 features), increasing model complexity. While some studies employ principal component analysis (PCA) for dimensionality reduction, their implementations often suffer from data leakage due to offline processing. Although leakage-free strategies do exist, their application requires caution—particularly in the presence of unit-to-unit variability. Such variability alters data distributions across units, underscoring the need for careful implementation and systematic comparison of different PCA variants.

The present study is deliberately structured around these shortcomings. First, an adaptive, per-unit standardization continuously updates mean and variance, thereby cancelling inter-unit drift—an innovation absent from earlier work (see Table 1). Second, the high-dimensional (201) AE features are compressed in real time by causal PCA variants trained only on data already available—either the full history of the training units or the measurements observed up to the current time step for the test unit. These causal schemes are benchmarked against a conventional PCA fitted on complete life-cycle data, which would introduce data leakage. This procedure reduces the size of the subsequent model without contaminating the training set with future information and preserves a high percentage of the data variance (e.g., more than 90%). Third, a two-stage architecture decouples data-rich spatial learning from data-scarce temporal refinement: a time-independent model (TIM) extracts an initial HI from every AE window, and its output conditions a time-dependent model (TDM) that captures degradation dynamics using the limited failure sequences available. Fourth, physics-informed Bayesian optimization (PBO) employs a dual-component loss comprising (a) explicit penalties that enforce Mo, Pr, and Tr, and (b) an implicit term that minimizes error relative to simulated proxy labels.

Rigorous leave-one-out cross-validation on twelve impacted, stiffened CFRP panels—each fold isolating both a validation unit and a test unit—yields a mean fitness of 90% (±2%). This performance surpasses earlier AE-based studies that relied on full life-cycle information or contained data-leakage concerns, establishing an HI for aerospace composites that is simultaneously monotonic, prognosable, and trendable under genuine real-time constraints while directly addressing the methodological gaps summarized in Table 1.

3. Experimental tests and feature extraction

Twelve composite skin-stiffener panels were subjected to compression-compression fatigue loading. 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 $[\pm 45/0/45/90/-45/0]_S$ and $[\pm 45/0/\pm 45]_S$, respectively. Various uncertainties, such as impact loading (representing phenomena like bird strikes or hail

Table 1
Health indicator modeling approaches for composite structures under fatigue loading. Note that only methods using SHM data are included. Impractical indices obtained by non-destructive methods not yet mature for in-situ monitoring (e.g., DIC or X-ray) or damage indices not quantifiable during operation (e.g., crack density or delamination size) are excluded.

Study	SHM Technique	Modeling Approach	Data Normalization	Physics-Based Constraints	Prognostic Criteria	Other Limitations
Liu et al. (2009)	AE; GW	**AE:** count feature of AE waveforms (data-driven); **GW:** energy of low-frequency signal from wavelet transform (data-driven)	max–min normalization	✘	Not reported	Unrealistic normalization using full data trajectories
Milanoski and Loutas (2021)	Strain (FBG)	Relative/cumulative strain (physics-based)		✘	Not reported	
Galanopoulos et al. (2021a)	Strain (FBG); AE	**FBG:** (1) relative strain (physics-based), (2) monotonicity-weighted fusion of relative strain from all FBGs (physics-based), (3) PCA-Q index from strain (data-driven); **AE:** windowed cumulative features (Hits & RA) (data-driven)		✘	Not reported	
Galanopoulos et al. (2021b)	Strain (FBG); AE	**FBG:** (1) relative strain, (2) relative/cumulative strain, (3) monotonicity-weighted fusion of relative strain, (4) monotonicity-weighted fusion of relative/cumulative strain (all physics-based); (5) PCA Euclidean distance, (6) PCA-Q index (both data-driven); **AE:** windowed cumulative features (Hits & RA) (data-driven)	PCA Euclidean distance normalized with radial basis function	✘	Qualitatively discussed	
Yue et al. (2022)	GW	Stiffness estimation from guided waves (physics-based)		✔ (implicit)	85%	Excitation frequency dependency; Mo not calculated using MMK metric
Moradi et al. (2023a)	AE	SSLSTM HI (intrinsically semi-supervised learning with embedded prognostic metrics) (data-driven)	Zero-mean normalization based on training data	✘	93% (±4.7%)	Test unit data used for validation during LOOCV, limiting real-world applicability
Galanopoulos et al. (2023)	Strain (FBG)	Fused strain-based HI via genetic algorithm (data-driven)		✔ (explicit)	87.5%	Trendability not considered; Mo not calculated using MMK metric
Moradi et al. (2024b)	GW	HT-SSCNN with ensemble learning (data-driven)		✔ (implicit)	93% (±5%) (Case 1); 81% (±20%) (Case 2)	Limited temporal resolution due to guided wave acquisition frequency
Current study	AE	PCA-TIM-TDM (spatiotemporal semi-supervised learning) with ensemble learning (data-driven)	Adaptive standardization	✔ (explicit)	90% (±2%)	Other limitations

impacting aircraft) and disbond (representing manufacturing defects), were considered in the experiments, making the HI construction task more challenging. Impact loading of 10 J was applied at various locations and times (see Fig. 1), and three panels also had disbond defects with different sizes intentionally introduced between the skin and T-stiffener during manufacturing (additional information can be found in Table 2). The panels underwent fatigue loading using an MTS machine at a frequency of 2 Hz and an R-ratio of 10.

Various techniques were employed to monitor the composite panels (Moradi et al., 2023a), (Zarouchas et al., 2021), including the AE method, which is the focus of this research. An amplitude threshold of 60 dB was set to capture relevant hits and filter out noise signals. Localization was achieved using the internal Vallen processor based on Geiger's model, while a filter excluded events with a position uncertainty greater than 50 mm. Six low-level features, namely amplitude, rise time, energy, counts, duration, and root mean square (RMS), were extracted from AE events. As can be seen in Fig. 2(a–e), following the necessary steps of windowing, signal processing, and feature extraction, a total of 201 statistical features (listed in Ref. (Moradi et al., 2023a).) were extracted for each time window. Both the low-level and statistical features are publicly available (Moradi et al., 2022b).

4. Evaluation of health indicators (HIs)

In this paper, two performance metrics are employed to evaluate the proposed scenario. The first evaluation metric is *Fitness*, which incorporates three physics-based prognostic criteria as outlined in the section "HIs' evaluation criteria." The second one is the root mean square error (RMSE) between the constructed HIs and the ideal simulated ones.

4.1. HIs' evaluation criteria

Given the crucial role of evaluation metrics for HIs in various stages of the proposed approach, it is important to first introduce them. The evaluation of a prognostic signature (i.e., HI) relies on three confirmed criteria (Mo, Pr, and Tr) (Moradi, 2024), (Moradi et al., 2024a), (Lei, 2016), (Saidi et al., 2017), (Coble, 2010), (Eleftheroglou et al., 2018) which are defined as follows:

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

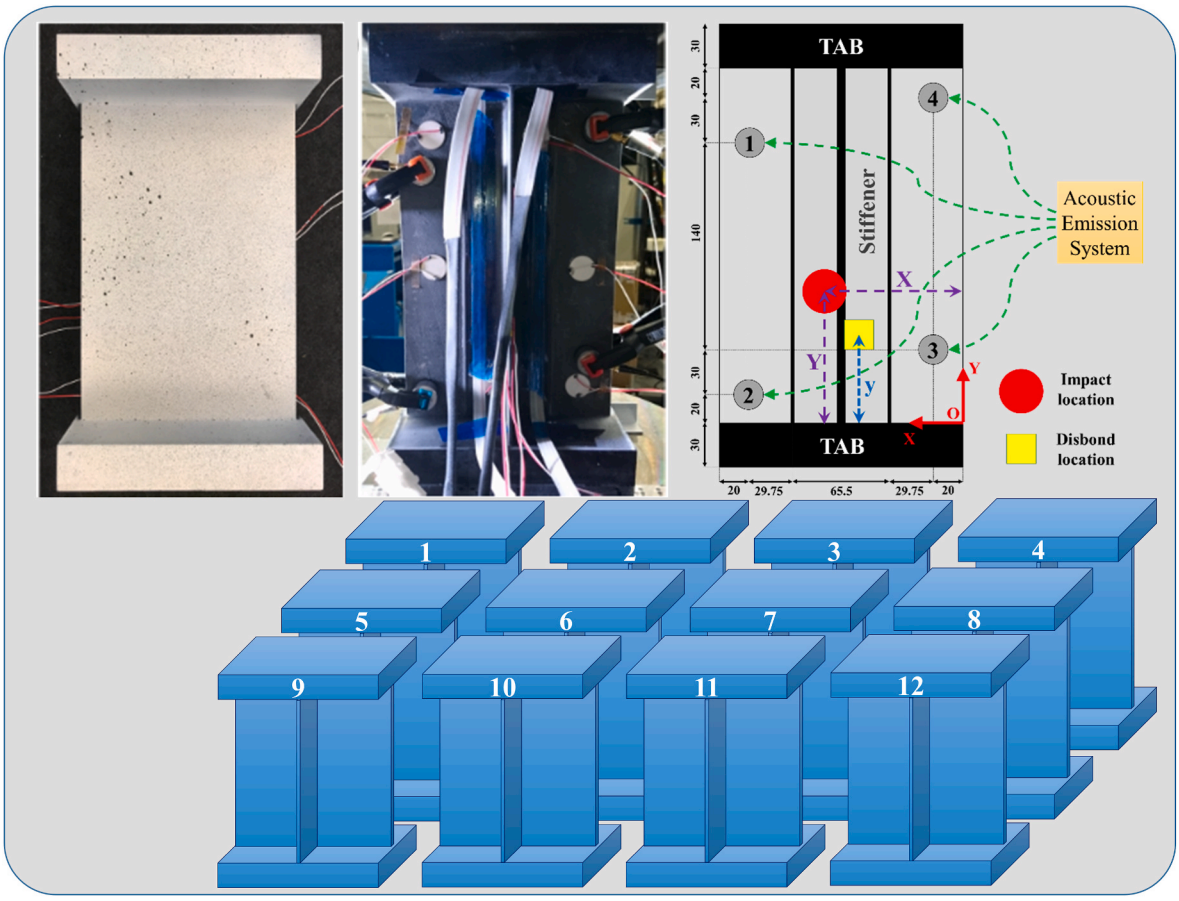

Fig. 1. Single T-stiffener panels. Left image: skin side; middle image: T-stiffener side; right image: acoustic emission sensor (gray circles), impact, and disbond locations with dimensions in [mm].

Table 2
The information of the single T-stiffener composite panels tested under run-to-failure compression-compression fatigue loading.

Composite Specimen	1	2	3	4	5	6	7	8	9	10	11	12
X-location of impact (mm)	50	115	82.5	50	50	50	50	50	115	50	50	115
Y-location of impact (mm)	80	160	140	160	160	160	160	160	80	160	160	80
Time of Impact (cycles)	0	0	0	5000	5000	5000	0	0	5000	5000	5000	5000
Size of disbond (mm)				15 × 20	20 × 20	20 × 25						
y-location of disbond (mm)				60	60	60						
Loading cycles	152,458	144,969	133,281	48,702	65,500	94,431	368,558	510,961	226,356	756,226	110,137	170,884

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

$$Pr = exp\left(-\frac{\sqrt{\frac{1}{M}\sum_{j=1}^{M}\left|x_j(N_j) - \left[\frac{1}{M}\sum_{i=1}^{M}x_i(N_i)\right]\right|^2}}{\frac{1}{M}\sum_{j=1}^{M}\left|x_j(1) - x_j(N_j)\right|}\right), \quad j = 1,2,\ldots,M \tag{3}$$

where $x(t_p)$ and $x(t_i)$ indicate HI at the times of t_p and t_i, respectively. $sgn(\cdot)$ is the sign (or signum) function. $cov(x_j, x_k)$ is the covariance, where x_j and x_k are the vector of HIs of the j^{th} and k^{th} specimens (among M composite specimens) that have N_j and N_k measurements, respectively. The standard deviations of x_j and x_k are represented by σ_{x_j} and σ_{x_k}, respectively. The evaluation metric chosen for Mo in Eq. (1) is the Modified Mann-Kendall (MMK). This metric, compared to the Sign and Mann-Kendall versions, is more robust to noise and takes into account the relationship between data points with time gaps larger than one unit (Moradi, 2024), (Moradi et al., 2023a), (Eleftheroglou et al., 2018). All three criteria (Mo, Pr, and Tr) are scored on a range from 0 to 1, with a score of 1 indicating the optimal performance of the HIs. Taking into account all of the aforementioned criteria, the Fitness metric is defined as follows:

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

Assuming that the control constants a, b, and c are equal to 1, the Fitness metric ranges from 0 (representing the minimum quality) to 3 (representing the maximum quality) for the evaluated HIs.

4.2. Deviation from the simulated ideal HIs

The second evaluation criterion is the RMSE, defined as:

$$RMSE = \sqrt{\frac{1}{N_j}\sum_{i=1}^{N_j}\left(T_i - HI_i^{(p)}\right)^2}, \quad j \in M, \ p \in \{1,2\} \tag{5}$$

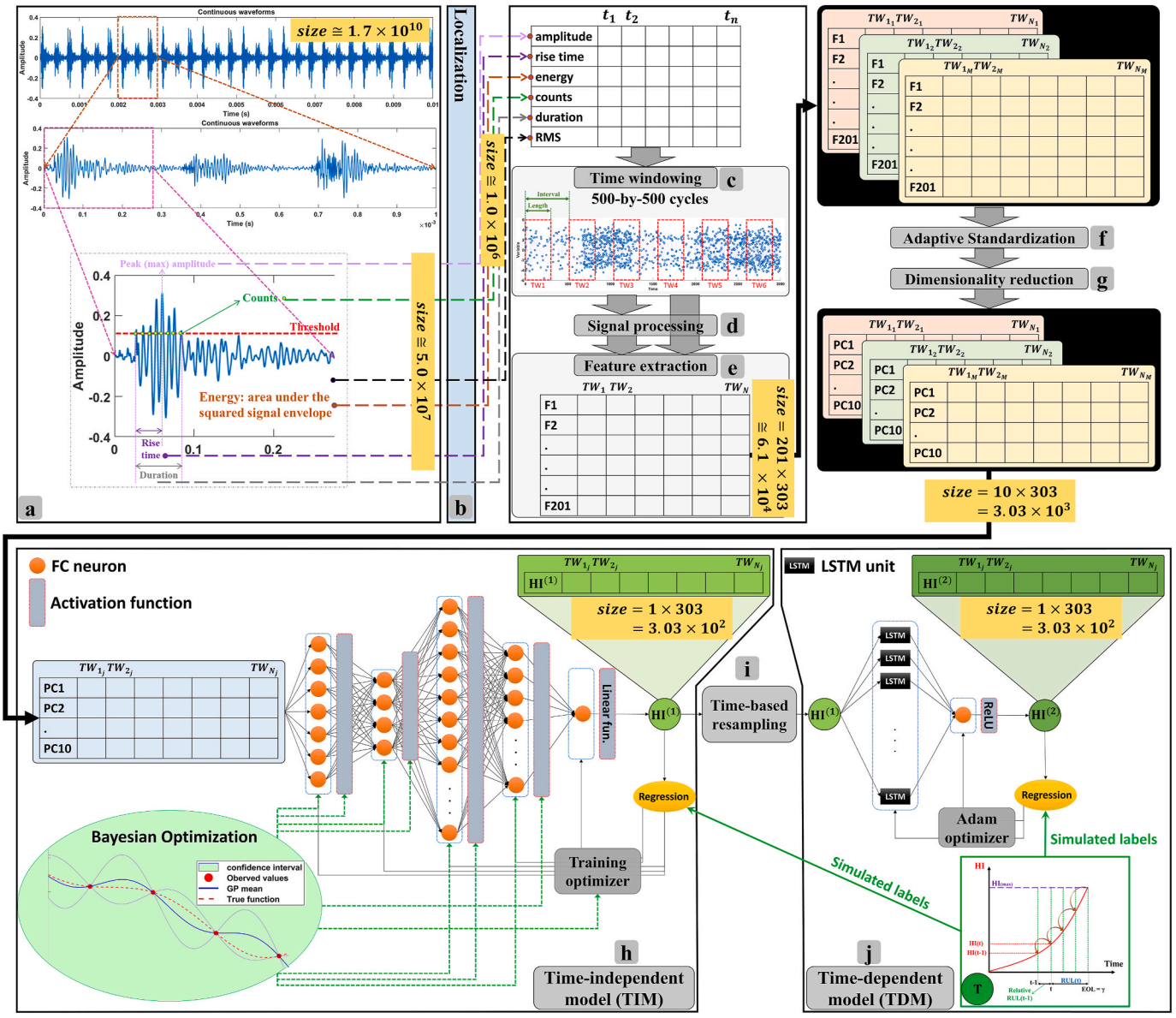

Fig. 2. The overall framework, from the illustrative SHM data to the final HI: (a) pre-processing (initial feature extraction), (b) localization, (c) windowing, (d) signal processing, (e) feature extraction, (f) adaptive standardization, (g) dimensionality reduction, (h) time-independent model, (i) time-based resampling, and (j) time-dependent model. Approximate data sizes shown at each stage correspond to the first specimen and are provided to illustrate the scale of data reduction achieved throughout the processing pipeline.

where N_j represents the length of the sequence ($HI^{(p)}$), T_i denotes the target value which is the simulated ideal HI under inductive SSL paradigm using the non-linear hypothetical function of $T_i = 100(t_i/t_{EoL})^2$ (for more details, refer to Ref. (Moradi, 2024), (Moradi et al., 2023a), (Moradi et al., 2023b)), and $HI_i^{(p)}$ is the network's output (either the 1st or 2nd level HI) at time step *i*. This metric provides a single score for the constructed HI of each unit (composite panel), among which the score of the test unit's HI (not including training or validation) will be primarily reported.

5. Methodology

The methodology mainly consists of two parts: a two-stage semi-supervised criteria-based fusion model and an ensemble model, which will be introduced in this section.

5.1. Two-stage semi-supervised criteria-based fusion model

The current research presents a machine learning (ML) approach that combines a dimensionality reduction (PCA-based technique), a time-independent (TIM), and a time-dependent (TDM) sub-model after up-sampling the time series in each batch. Since there are no true values available for the HIs, an inductive SSL paradigm is employed to first simulate labels as the ideal HIs (Moradi et al., 2023a), where the prognostic metrics (Mo, Pr, and Tr) are implicitly implemented and the given EoL information is leveraged (Moradi et al., 2023b). This framework falls under the category of inductive learning algorithms known as intrinsically SS (van Engelen and Hoos, 2020). These algorithms are improvements over traditional supervised algorithms as they allow labeled and unlabeled data to be directly used for optimizing an objective function with multiple components. The overall framework, as depicted in Fig. 2, encompasses the entire process, from raw SHM data to the final HI (the 2nd level HI), including the new proposed approach

(stages f to j). The remaining steps of the framework, from adaptive standardization (stage f) through TDM (stage j), are provided in this section as the primary contributions of the current work. It is important to mention that the units (composite panels) are divided into three portions: training (D), validation (v), and test (τ) sets. Additionally, the training data of D itself is further split into training (D') and validation (v') subsets specifically for TIM. This division strategy ensures that TIM is adequately trained and validated before its application to the overall HI construction process.

5.1.1. Adaptive standardization

The normalization (referring to max-min normalization) or standardization (referring to zero-mean normalization) are commonly applied pre-processing techniques for inputs before being fed into a model. These advantageous pre-processing techniques are also employed between the hidden layers of an artificial neural network (ANN) model. However, they should be utilized with caution in the field of prognostics or other fields with a similar objective because future data is not accessible to the model, especially during its testing phase. In fact, whether the input data are normalized or standardized given the entire trajectory, the input data distribution is already placed into a known range. In this manner, the problem has somehow been turned into an interpolation task for the data-driven model when it is in fact an extrapolation task, which is more difficult for ML models. This point is crucial for assessing prognostic-relevant methods, such as HI construction or RUL prediction models.

According to the explanations above, the plausible standardization technique can be used for both training and test data based on the mean value and standard deviation derived exclusively from the training data. However, because we employed the same technique and noticed its misleading impacts on results (Moradi et al., 2023a), this may not be useful for new unseen data (validation or test portions). As a result, an adaptive standardizing technique is being developed. Assuming that μ_i and σ_i are the mean value and standard deviation of the data ($x_{1:i}$) up to the present (time step t_i), the data are standardized as follows:

$$x_i^{st} = \frac{x_i - \frac{\sum_{j=1}^{i-1} x_j^{st} + x_i}{i}}{\sqrt{\frac{\sum_{j=1}^{i-1}\left(x_j^{st} - \frac{\sum_{j=1}^{i-1} x_j^{st} + x_i}{i}\right)^2 + \left(x_i - \frac{\sum_{j=1}^{i-1} x_j^{st} + x_i}{i}\right)}{i}}} = \frac{x_i - \mu_i}{\sigma_i} \tag{6}$$

$$\mu_i = \frac{\sum_{j=1}^{i-1} x_j^{st} + x_i}{i} \tag{7}$$

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{i-1}\left(x_j^{st} - \mu_i\right)^2 + (x_i - \mu_i)}{i}} \tag{8}$$

It is important to note that the data preceding the current time step ($x_{1:i-1}$) is not standardized at the current time step (t_i), but has been standardized in previous time steps (t_1, t_2, …, and t_{i-1}). Only the current data (x_i) undergoes the standardization process at the current time step (t_i), rather than the preceding data. This procedure is carried out for the extracted AE features of each unit independently, which is acceptable and applicable from the prognostics standpoint.

5.1.2. Dimensionality reduction

Due to the large number of features obtained from AE data, the subsequent models become more complex. To address this issue, it is beneficial to reduce the feature dimensionality. To accomplish this, PCA-based methods are employed (de Moura et al., 2011), (Dunteman, 1989), (He et al., 2021), (O. P and Van Der Maaten, 2009). The aim is to decrease the number of features from 201 statistical features to 10 principal components (PCs). While additional PCs could be extracted and used as inputs for subsequent models, it is deemed acceptable to retain the reconstructed variance (∼90%) of AE features in order to maintain simplicity in the following models.

The application of PCA in the context of prognostics poses a challenge, as it is essential to avoid utilizing future data during the decomposition of eigenvectors (*coefficients*), especially when testing the model. In this paper, we explore two primary approaches to applying PCA:

Approach A: The coefficients of PCs are extracted solely from the training units. Subsequently, these coefficients are utilized to construct PCs for the test units. There are two variations in this approach: **A.1**, where all data from the training units is concatenated (making a bigger matrix), and the coefficients are then extracted; and **A.2**, where coefficients are individually extracted from each training unit, and the mean value of the coefficients is used for testing.

Approach B: Coefficients are independently extracted for each unit, regardless of whether they are from the training or test specimens. However, this approach is subject to two limitations. Firstly, the number of time steps during which measurements are collected must exceed the required number of PCs, due to matrix size requirements. Secondly, when considering the overall prognostics problem for each unit individually, the usage of future data (beyond the current time step) is prohibited, particularly during the test phase, due to the unavailability of future SHM measurements in practical scenarios.

Based on the authors' knowledge, approach B of estimating PCs yields higher performance when all data from the beginning until the EoL is considered. However, as mentioned earlier, the use of future data is impossible for prognostics. To address time-dependent data, several extensions to PCA methods have been introduced, including dynamic PCA (DPCA), recursive PCA (RPCA), and moving-window PCA (MWPCA) (Barabadi et al., 2011). While DPCA was developed to handle autocorrelation, RPCA and MWPCA are more suitable for dealing with nonstationary data. Among them, MWPCA forgets the older data (before window) entirely, which is not appropriate for the current prognostics purpose, especially when using passive SHM methods like acoustic emission. To overcome this limitation, RPCA can be employed to incorporate all historical SHM data from the beginning of monitoring up to the current time. The approach B with all historical data till EoL is denoted as code **B.1** (which is not acceptable in prognostics) and RPCA is coded as **B.2**, which is described in more detail in the following. As the dimensionality reduction stage inside the suggested framework, all approaches A.1, A.2, B.1, and B.2 are used and compared in this work. The implementation and differences among these four PCA-based types are illustrated in Fig. 3. Although B.1 is not feasible in real-world applications, it is still used to demonstrate false positive performance and compare it with others.

In RPCA, assuming that the mean and covariance of all adaptively standardized observations $x_{1:i}^{st}$ up to the present (time step t_i) are $\widehat{\mu}_i$ and S_i, respectively, the updated mean $\widehat{\mu}_{i+1}$ and the updated covariance matrix S_{i+1} given the new adaptively standardized observation x_{i+1}^{st} (at time step t_{i+1}) can be calculated by:

$$\widehat{\mu}_{i+1} = \left(1 - \frac{i}{i+1}\eta\right) x_{i+1}^{st} + \frac{i}{i+1}\eta\widehat{\mu}_i \tag{9}$$

$$S_{i+1} = \left(1 - \frac{i}{i+1}\eta\right)\left(x_{i+1}^{st} - \widehat{\mu}_{i+1}\right)\left(x_{i+1}^{st} - \widehat{\mu}_{i+1}\right)^T + \frac{i}{i+1}\eta S_i \tag{10}$$

where $0 \le \eta \le 1$ is a forgetting factor. In the context of a geometric progression, this is comparable to calculating a weighted mean and covariance of $x_{1:i+1}^{st}$, where earlier values are down weighted exponentially. Older observations in RPCA are faster forgotten when using a

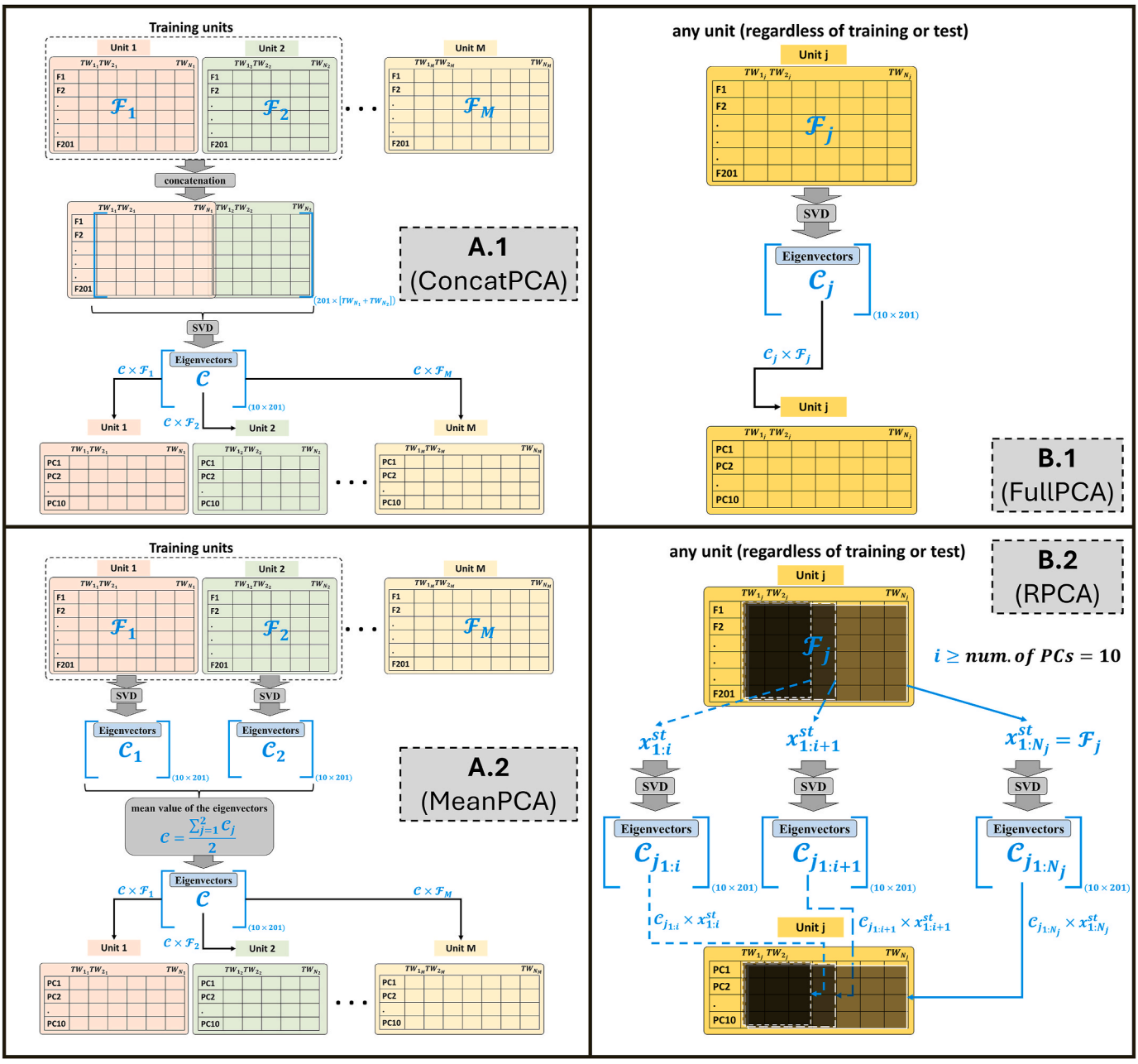

Fig. 3. Illustration of the implementation and differences among the four PCA-based techniques (A.1, A.2, B.1, and B.2). $\mathscr{T}_j$ represents the full matrix of features for the j-th unit, and $x^{st}_{1:i}$ is the segmented matrix of features from the beginning (the first time step) up to the present (time step t_i). $\mathscr{C}$ denotes the matrix of eigenvectors (coefficients), and $\mathscr{C}_j$ is the matrix of eigenvectors for the j-th unit, calculated using singular value decomposition (SVD). In this illustration, 10 principal components (PCs) are targeted.

forgetting factor of $\eta < 1$, whereas they are slower forgotten by applying a forgetting factor of $\eta \to 1$. If it is desired to maintain all prior observations, which is appropriate for the current study, $\eta = 1$ should be adopted. By adopting RPCA, the proposed approach retains adaptability and compatibility with new incoming data while still maintaining the integrity of historical information critical for prognostics applications.

5.1.3. Time-independent model (TIM)

In this section, we establish an initial neural network architecture through a trial and error process. Subsequently, we employ the PBO algorithm to optimize the relevant constructive hyperparameters and network architecture, as illustrated in Fig. 4. Since only ten features (PCs) remain after dimensionality reduction, a deep neural network (DNN) with a few layers serves as a suitable starting point for the architecture designated to the regression task, fitting the 10 PCs to a value (HI). Detailed information about the initial network, its loss function, hyperparameters, etc., can be found in subsection **a**, while further details on the PBO and its objective function are provided in subsection **b**.

a. Training and architecture of TIM

A DNN with four layers, including a linear transfer function as the output layer, is designed to fit the 10 PCs to the ideal simulated HI. The loss function used for comparing predictions and targets is a modified mean absolute error (MMAE), given by Eq. (11):

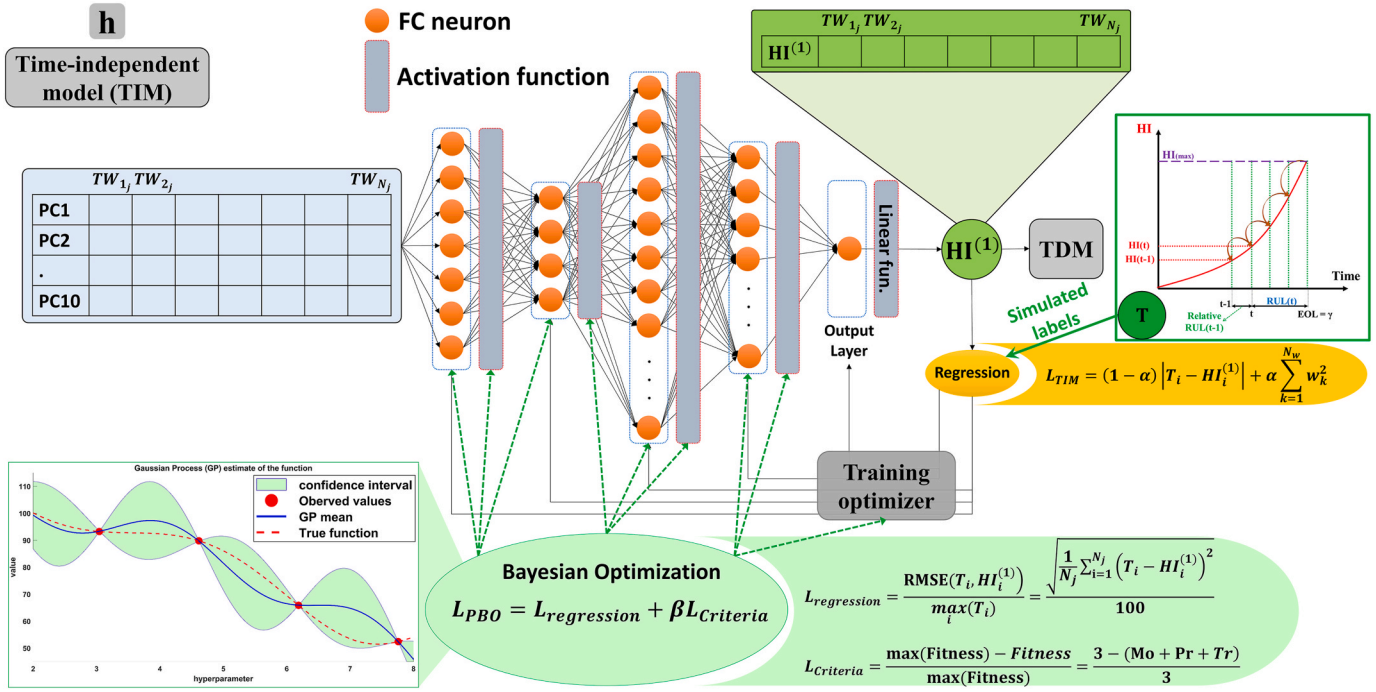

Fig. 4. Time-independent model (TIM) fed by ten principal components (PCs) and yielding the 1st level HI.

$$L_{TIM} = (1-\alpha)\frac{1}{R}\sum_{i=1}^{R}\left|T_i - HI_i^{(1)}\right| + \alpha\sum_{k=1}^{N_w} w_k^2, \quad i \in D \tag{11}$$

where R represents the number of responses chosen among the training data points ($R \subset [1,N_1] \cup [1,N_2] \cup \ldots [1,N_j] \ldots \cup [1,N_M]$, N_j representing the number of responses for the j[th] unit and M representing the number of units). T_i and $HI_i^{(1)}$ denote target value and the network's output for response i, respectively. N_w denotes the number of learnable weights of TIM. The regularization parameter α is introduced to enhance generalization by modifying the performance function. The use of this performance function reinforces the ANN to have smaller weights and biases, resulting in smoother responses and reducing overfitting. For the current work, α is set to 1.

The DNN model is trained and validated using only the training composite panels data ($i \in D$). 30% of the training data is reserved for validation purposes ($v' \subset D$). While the maximum number of training epochs is set to 1000, the DNN's output is based on the best validation loss, with the validation check patience set to 10. According to the optimizers and their default values in the MATLAB R2022a framework, the remaining hyperparameters are determined through the utilization of the PBO algorithm, which will be further elaborated on in the following section.

b. Physics-based Bayesian optimization (PBO)

The hyperparameters optimized by the BO algorithm (Snoek et al., 2012) include training optimizer algorithms, the number of neurons in each fully connected (FC) layer, and the activation function for all hidden layers. The TIM's weights and biases can be trained using a variety of optimizers. The three types of optimizer algorithms Levenberg-Marquardt (LM) (Hagan et al., 1997), (Hagan and Menhaj, 1994), (Marquardt, 1963), Bayesian regularization (BR) (Foresee and Hagan, 1997), (MacKay, 1992), and resilient backpropagation (RB) (Hagan et al., 1997), (Riedmiller and Braun, 1993) are considered as the search space of the first optimizable variable that can be tuned by BO. The number of neurons allocated for FC layers 1, 2, 3, and 4 are within the ranges [1,50], [1, 50], [1, 50], and [1, 10], respectively. The activation function is the last optimizable variable, and it is assigned the same type for all hidden layers. It is selected from a categorical space according to MATLAB terminology. The aforementioned hyperparameters and their respective search spaces are provided in Table 3.

A new objective function inspired by the physics of the degradation and health management for the BO algorithm is introduced, which considers the evaluation metrics of the HI. Accordingly, the BO algorithm will hereinafter be referred to as "Physics-based BO (PBO)". The PBO objective function comprises two components: regression and criteria loss functions. The regression loss function is based on RMSE between the SS-based simulated labels and predictions, calculated solely on the validation (composite panels) set ($i \in v$). The criteria loss function includes Mo, Pr, and Tr, which are computed using all datasets, including both training and validation portions ($i \in v \cup D$). Since the BO algorithm does not initially have a clue about the optimal solutions' space, this can lead to a wide range of responses. It results in slower convergence and increased time consumption. The regression loss function serves as a guide for the model by providing a general pattern. On the other hand, the criteria loss function strengthens the significance of prognostic metrics to prevent the algorithm from merely adhering to the simulated labels, allowing it to explore other viable and meaningful solutions. This strategy drives the PBO algorithm to seek out diverse and potentially superior solutions, improving its capacity to adapt and converge towards better outcomes.

The equations of the regression and physics-based criteria loss functions that constitute the PBO objective function are as follows:

$$L_{\text{reg}} = \frac{\text{RMSE}\left(T_i, HI_i^{(1)}\right)}{\max_i(T_i)} = \frac{\sqrt{\frac{1}{N_j}\sum_{i=1}^{N_j}\left(T_i - HI_i^{(1)}\right)^2}}{100}, \quad i \in v, \ j \in M \tag{12}$$

$$L_{Criteria} = \frac{\max(Fitness) - Fitness}{\max(Fitness)} = \frac{3 - (Mo + Pr + Tr)}{3}, \quad i \in (v \cup D) \tag{13}$$

$$L_{PBO} = L_{reg} + \beta L_{Criteria} \tag{14}$$

The parameter β determines the significance of $L_{Criteria}$ in relation to L_{reg}. $L_{Criteria}$ is normalized using the maximum Fitness score, resulting in a

Table 3
The hyperparameters of TIM and their search spaces for optimization by PBO.

Hyperparameter	Search space		Type
Optimizer	Levenberg-Marquardt (LM) Bayesian regularization (BR) Resilient backpropagation (RB)		categorical
Number of neurons at FC layer 1	[1, 50]		integer
Number of neurons at FC layer 2	[1, 50]		integer
Number of neurons at FC layer 3	[1, 50]		integer
Number of neurons at FC layer 4	[1, 10]		integer
	Name	***function***	
Activation functions	Linear	$\sigma(z) = z$	categorical
	Rectified linear units (ReLU)	$\sigma(z) = \begin{cases} 0, & z \le 0 \\ z, & z > 0 \end{cases}$	
	Saturating linear	$\sigma(z) = \begin{cases} 0, & z \le 0 \\ z, & 0 \le z \le 1 \\ 1, & z \ge 1 \end{cases}$	
	Symmetric saturating linear	$\sigma(z) = \begin{cases} -1, & z \le -1 \\ z, & -1 \le z \le 1 \\ 1, & z \ge 1 \end{cases}$	
	Hard-limit	$\sigma(z) = \begin{cases} 0, & z < 0 \\ 1, & z \ge 0 \end{cases}$	
	Symmetric hard-limit	$\sigma(z) = \begin{cases} -1, & z < 0 \\ 1, & z \ge 0 \end{cases}$	
	Log-sigmoid	$\sigma(z) = \frac{1}{1+\exp(-z)}$	
	Hyperbolic tangent sigmoid	$\sigma(z) = \frac{2}{1+\exp(-2z)} - 1$	
	Elliot symmetric sigmoid	$\sigma(z) = \frac{z}{1+\lvert z\rvert}$	
	Radial basis	$\sigma(z) = \exp\left(-z^2\right)$	
	Normalized radial basis	$\sigma(z) = \frac{\exp\left(-z_i^2\right)}{\sum_{j=1}^{J}\exp\left(-z_j^2\right)},\ i = 1,2,3,\ldots,J$	
	Softmax	$\sigma(z) = \frac{\exp(z_i)}{\sum_{j=1}^{J}\exp(z_j)},\ i = 1,2,3,\ldots,J$	
	Triangular basis	$\sigma(z) = \begin{cases} 0, & z < -1 \\ 1-\lvert z\rvert, & -1 \le z \le 1 \\ 0, & z > 1 \end{cases}$	
	Inverse	$\sigma(z) = \frac{1}{z}$	
	Competitive	$\sigma(z) = \begin{cases} 1, & j = \underset{i}{\operatorname{argmax}}\,(z_i), \quad i = 1,\ldots,J \\ 0, & j \ne \underset{i}{\operatorname{argmax}}\,(z_i), \quad i = 1,\ldots,J \end{cases}$	

range of [0, 1]. On the other hand, L_{reg} is normalized based on the maximum target value, which is 100. It should be noted that the ideal HI values are simulated within a range from 0 (representing a healthy state) to 100 (indicating a failure state). The PBO algorithm was given 100 trials with an exploration ratio of 0.8 in parallel computing to optimize the hyperparameters.

5.1.4. Time-based resampling

After the TIM step, the 1st level predicted HI can serve as a prognostic parameter to be imported into a prognostic model for predicting RUL. Despite the undeniable time dependence between the SHM data according to the physics of fatigue phenomenon, progressive damage, and the AE method, this aspect has been overlooked so far in the proposed approach. Prior to designing TDM, the input data needs to undergo resampling to ensure that all sequences of input $HI^{(1)}$ within a batch have the same length. Conventional padding techniques like zero padding are unsuitable in this context, as the HI values with respect to the percentage of lifetime should exhibit similarities. For example, if the batch size is 2 and the lengths of the two intended HIs are 100 and 1000, extending the first HI with 900 zero values to match the length of the second HI would incorrectly set the HI at the EoL for the first unit as 0 instead of its actual (compatible) value, which should be 100, as in the second HI. Likewise, conventional interpolation methods are not viable due to the non-linear correlation between the number of data points in HI and the EoL. Due to the varying recording rate of the passive AE system depending on the predetermined amplitude threshold value and uncertain progressive damage, the length (number of data points) of HI for a unit with a bigger EoL may be shorter than that with a smaller EoL.

To address these challenges, the current study adopts a technique for up-sampling referred to as "time-based resampling" (see Fig. 5). The first step is to convert the time vectors of HIs into percent lifespans, with values in the range of [0%, 100%]. The shorter HI vectors (in terms of the quantity of data points) in each batch are then up-sampled (based on the associated time vectors) to match the length of the longest HI vector. For every batch separately, this procedure is performed. It is crucial to remember that in such a scenario, the batch size cannot be the same as the total number of training units. This is due to the TIM's tendency to just learn the position of data, disregarding its value, leading it to predict from 0 (healthy) at the beginning to 100 (failure) at the EoL based exclusively on the position of the incoming data. For instance, if HIs of equal lengths of 1000 are employed, the TIM model would learn that, regardless of the input value, position 1 should result in a zero value and position 1000 should result in a hundred. As the test set should not logically be resampled, training a model in this way would result in much worse performance on the test set. As a result, the batch size should be less than the number of all the training units. In the current investigation, 10 composite panels are kept for training, and one panel is designated for testing and another for validation in all scenarios. Consequently, a batch size of 2 or 5 appears more reasonable, with 2

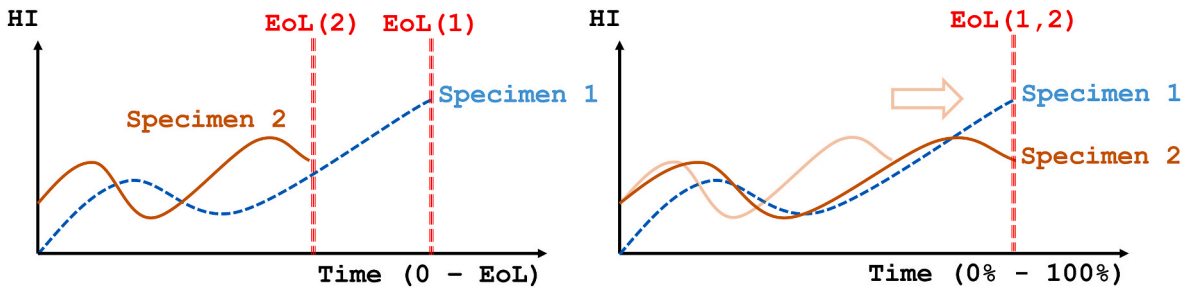

Fig. 5. Sequence length synchronization and time-based resampling.

being the preferred choice. Lastly, it should be noted that if a batch size of 1 is chosen, time-based resampling becomes irrelevant.

5.1.5. *Time-dependent model (TDM)*

In this subsection, a sequence-to-sequence (seq2seq) regression model called TDM is introduced to address the time dependence among AE data. Prior to this, using the previous regression model, TIM, 10 high-level extracted features (PCs) are mapped to the 1st level HI without considering the time dependence among AE data. The TDM model takes into account the temporal relationship between data points, which is crucial for accurate prognostic applications.

To maintain a long-term record of sequential inputs, the long short-term memory (LSTM) layer (Hochreiter and Schmidhuber, 1997) is a suitable candidate and serves as a key component in the TDM (see Fig. 6). The LSTM layer consists of 10 units, initialized with zero values for hidden and cell states. The pseudo-Huber loss function, which is a smooth approximation of the Huber loss function (Huber, 1992), is used for the seq2seq regression task (Riedmiller and Braun, 1993), (Hartley and Zisserman, 2004):

$$L_{TDM} = \frac{1}{N_j}\sum_{i=1}^{N_j}\delta^2\left(\sqrt{1+\left(\frac{T_i - HI_i^{(2)}}{\delta}\right)^2}-1\right),\quad i\in D,\ j\in M \tag{15}$$

Here, δ controls the steepness at extreme values and is set to 20 in our work after trial and error (20% of the maximum target value, considering that $\max_i(T_i) = 100$). An Adam optimizer (Kingma and Ba, 2015) is employed to train the TDM, with an initial learning rate of 0.01, a learning rate drop factor of 0.1, a learning rate drop period of 10, and a gradient threshold of 1, all determined through trial and error. The network's output is based on the best validation loss ($j \in v$), with the validation check frequency set to 50 iterations (the number of learned batches) and the validation check patience set to 50, despite the fact that the maximum number of training epochs was set to 2000. Since the sequences in each batch are already identical in length, a batch size of 2 was chosen, as described in the preceding section.

5.2. *Ensemble models*

After constructing HIs based on individual TIM-TDM models, an ensemble learning (EL) technique can be employed to create a meta-model capable of addressing inherent randomness in ML models and uncertainties in model structure. This is particularly crucial when the sample size is limited, as in this study. Combining various single prediction models into an EL model (ELM) can effectively leverage the strengths of different base models, thereby enhancing reliability and accuracy (Opitz and Maclin, 1999), (Polikar, 2006). There are three main categories of EL techniques: bagging (Breiman, 1996), boosting (Friedman et al., 2000), and stacking (Wolpert, 1992). Bagging (e.g., random forest (RF) (Breiman, 2001)) involves bootstrapping (random sampling) and aggregation (averaging the base learners' outputs). Boosting (e.g., adaptive boosting (AdaBoost) (Freund and Schapire, 1995) and extreme gradient boosting (XGB) (Chen and Guestrin, 2016)) combines sequentially arranged base learners. In contrast to the former

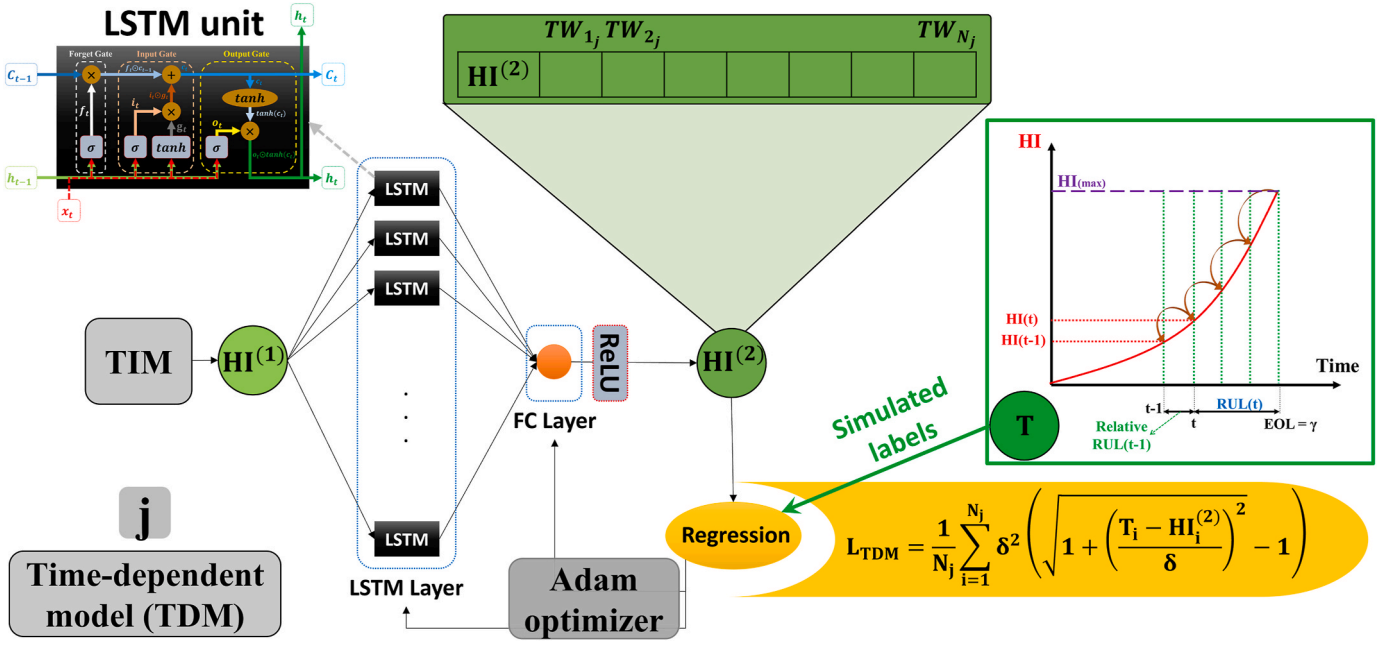

Fig. 6. Time-dependent model (TDM) fed by the 1st level HI and yielding the 2nd level HI.

methods, which use homogenous base models, stacking utilizes heterogeneous base learners and integrates them through training a meta-model (blender).

From another perspective, ELMs based on averaging can be broadly categorized into two types: simple averaging ensemble (SAE) (Haykin, 1998) and weighted averaging ensemble (WAE) (Hashem, 1997), with the former being a specific case of the latter. In SAE, the ELM combines predictions from multiple base models. However, a limitation is that each model contributes equally to the ensemble prediction, regardless of its performance. In contrast, WAE assigns weights to ensemble members based on trust in their predictive capabilities, allowing for a more sophisticated contribution.

Although several ELMs (such as RF, least squares linear, support vector machine (SVM), boosting (LSBoost), Bayesian linear regression, Gaussian Process (GP) regression, and Gaussian kernel regression) were explored, the presented approach focuses exclusively on averaging ELMs utilizing distinct weighting techniques. This choice is due to the superior results in our case after trial, along with the simplicity and efficiency of these models compared to others.

Initially, the approach employs LOOCV (Hastie et al., 2001), where a single unit (composite panel) is designated for testing, leaving 11 specimens for subsequent processing. Among these, one unit serves as validation, while the remainder (10 units) form the training set. This validation unit can be interchanged within the 11, resulting in 11 distinct combinations (subsets). For each combination, the TIM-TDM algorithm is individually trained ten times, each time using a different random seed number (Matsumoto and Nishimura, 1998), (Marsland, 2011). This leads to the generation of ten diversely trained models. Consequently, a total of 110 predictions ($HI^{(2)}$) are derived, accounting for both the 10 repetitions and the 11 combinations. Subsequently, a process involving SAE and WAE is employed to ensemble the 110 HIs.

The general WAE can be expressed as follows:

$$f_{WAE} = \sum_{k=1}^{K} \overline{\omega}_k f_k \tag{16}$$

where f_k represents the k^{th} individual TIM-TDM model and $\overline{\omega}_k$ is its normalized influential weight:

$$\overline{\omega}_k = \frac{\omega_k}{\sum_{k=1}^{K} \omega_k} \tag{17}$$

Here, ω_k represents the weight of the k^{th} individual model, which can be determined based on various error metrics. These metrics encompass the mean square deviation (MSE) between the predictions of the k^{th} model ($HI^{k(2)}$, (2) representing the 2^{nd} level HI) and the simulated HIs (T^k), the model's Fitness (prognostic criteria), or a combination of both metrics:

$$\omega_k^{MSE} = \frac{1}{MSE\left(T^k, HI^{k(2)}\right)} = \frac{1}{\frac{1}{|\nu \cup D|}\sum_{j\in(\nu\cup D)}\left[\frac{1}{N_j}\sum_{i=1}^{N_j}\left(T_i^k - HI_i^{k(2)}\right)^2\right]}, \; k \in (\nu \cup D) \tag{18}$$

$$\omega_k^{Fitness} = a.Mo_{HI} + b.Pr_{HI} + c.Tr_{HI}, \;\; k \in (\nu \cup D) \tag{19}$$

$$\omega_k^{Fitness\_MSE} = \frac{\omega_k^{Fitness} - \min_j\left(\omega_j^{Fitness}\right)}{\max_j\left(\omega_j^{Fitness}\right) - \min_j\left(\omega_j^{Fitness}\right)} + \frac{\omega_k^{MSE} - \min_j\left(\omega_j^{MSE}\right)}{\max_j\left(\omega_j^{MSE}\right) - \min_j\left(\omega_j^{MSE}\right)}, \; k,j \in (\nu \cup D) \tag{20}$$

where $|\nu \cup D|$ is the size of the set (which is 11). In the combined form Eq. (20), each weight is first scaled in a range [0, 1] to avoid compromising the effectiveness of its scale. If all $\overline{\omega}_k$ values are uniformly set to one, the SAE approach is adopted. After computing weights using Eqs. (18)–(20), one can find a subset of superior models by using the weight rankings to increase efficiency and guarantee improved result robustness. The remaining weights are then reset to zero. In this work, the top 10 base learners are kept using Eqs. (18) and (19), and afterwards, $\omega_k^{Fitness\_MSE}$ and $\overline{\omega}_k$ are calculated.

Unlike the fixed architecture of TDM, it is important to highlight that the TIM component of the base models, involving architecture optimization through PBO, dictates whether the base learners are homogeneous or heterogeneous. When TIM is executed individually for each validation index-based combination, the resulting TIM architectures, optimizers, and other hyperparameters are likely to differ. However, to mitigate resource-intensive processes, TIM's hyperparameters are fine-tuned using PBO solely for the first validation index-based combination in the set ($\nu \cup D$). For the remaining combinations, the same hyperparameters are retained, resulting in homogeneous TIMs across the 110 base learners.

6. Results and discussion

To comprehensively assess various combinations and ascertain the effectiveness, validity, and stability of the proposed framework, all 12 folds were examined. Within each fold, a single stiffener composite panel was designated as the test set, while the remaining panels served as training and validation data. Since there are 11 alternatives that can be used as validation and this choice is effective, this process was repeated across 11 different validation index-based combinations, iterated 10 times using distinct random seed numbers. Essentially, this approach emulates the LOOCV methodology, wherein a holdout validation strategy (Arlot and Celisse, 2010) is adopted within each fold. This strategy affirms the evaluation of the model's generalizability. These 1320 runs ($12 \times 11 \times 10$) were conducted for all four types of PCA-based data reduction techniques to facilitate a comprehensive comparison.

The implementation of the DL framework and signal processing steps was carried out using MATLAB R2022a. The aforementioned runs were executed utilizing a high-performance computing cluster (Beowulf style) with 20 processors on a single node. In this section, we first present the results of the proposed methodology up to the ensemble stage, termed Base learner models. Subsequently, we delve into the outcomes of the Ensemble learner models. Finally, we discuss and compare the results in comparison to the state-of-the-art works in the field.

6.1. Base learner models

Fig. 7 displays the simulated ideal HIs alongside a selection of promising candidates from the outcomes of $HI^{(1)}$s (TIM outputs) and $HI^{(2)}$s (TDM outputs) for different PCA-based techniques. Each case represents a holdout validation scenario. The error depicted in Fig. 7 illustrates the RMSE between the simulated ideal HIs and the constructed HIs. Notably, $HI^{(1)}$s generated by TIM exhibit significant fluctuations, whereas TDM yields smoother $HI^{(2)}$s. There have been some noteworthy observations, highlighting the limitations of TIM, such as cases where $HI^{(1)}$s exhibit a declining trend. On the other hand, TDM successfully corrects such patterns. Concerning the behavior of $HI^{(2)}$s, multiple incremental steps were observed over the fatigue life in many instances (e.g., A.1 in Fig. 7). These steps can potentially signify distinct damage states in composite panels, offering insights for subsequent prognostic models in RUL prediction. However, linking these steps with physical damage states in a stable and meaningful manner necessitates substantial effort and future experiments.

Table 4 presents the evaluation metrics for the HIs shown in Fig. 7. Notably, due to its time-dependent nature, all scores for $HI^{(2)}$s surpass those for $HI^{(1)}$s. The proposed model not only provides a streamlined

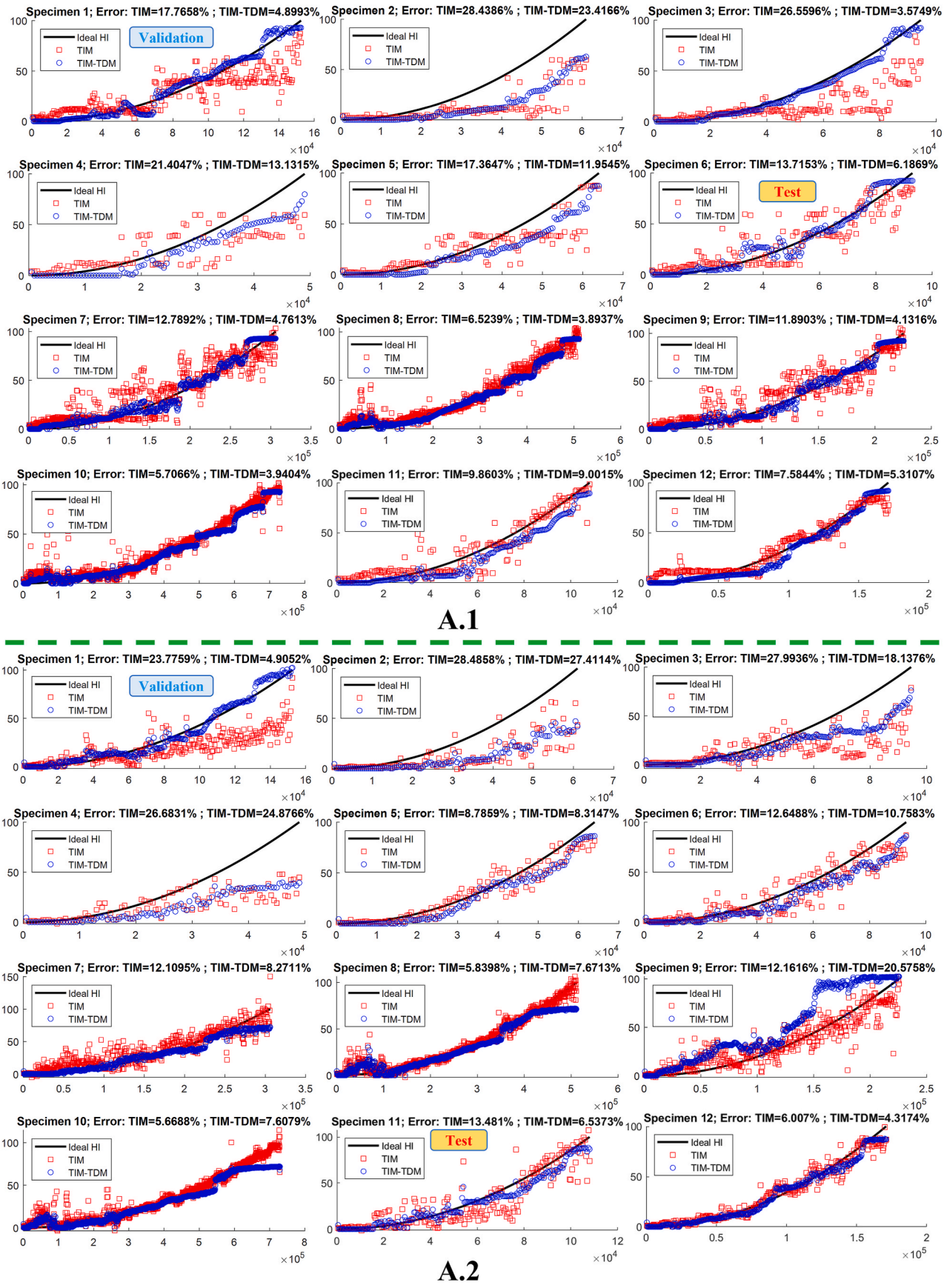

Fig. 7. Qualified candidates for $HI^{(1)}$s (TIM outputs) and $HI^{(2)}$s (TDM outputs) for different PCA-based techniques (A.1, A.2, B.1, and B.2). *(continued on next page)*

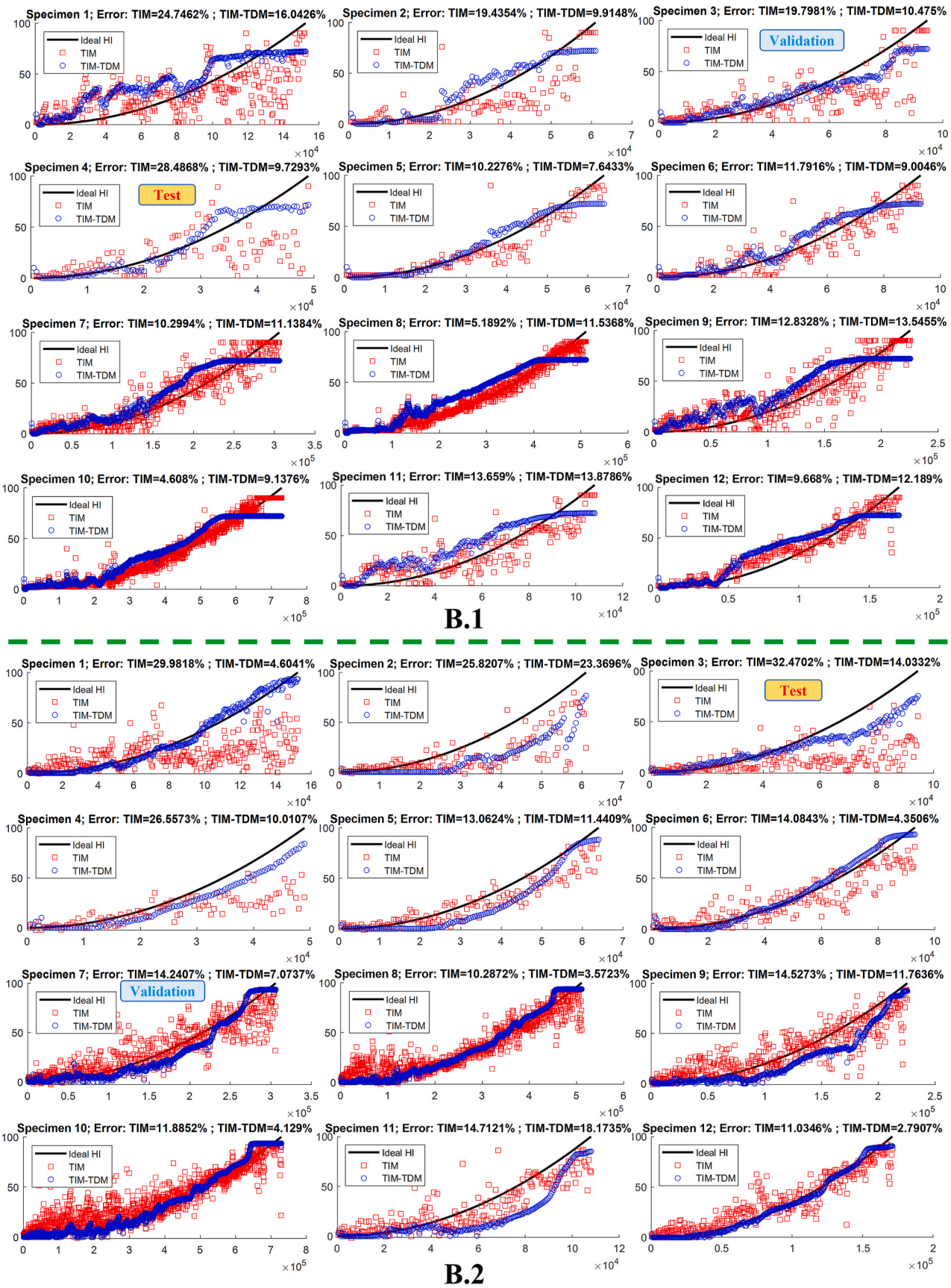

Fig. 7. (*continued*).

Table 4
The desirable candidate outcomes of each model run individually for various PCA methods.

PCA type	Test specimen	Validation specimen	Mo		Pr		Tr		Fitness		RMSE of test	
			TIM	TDM	TIM	TDM	TIM	TDM	TIM	TDM	TIM	TDM
A.1	6	1	0,98	1,00	0,82	0,90	0,66	0,94	2,46	2,84	13,72	6,19
A.2	11	1	0,99	1,00	0,72	0,76	0,63	0,92	2,34	2,68	13,48	6,54
B.1	4	3	0,98	1,00	0,92	1,00	0,43	0,91	2,32	2,90	28,49	9,73
B.2	3	7	0,95	1,00	0,71	0,92	0,40	0,90	2,06	2,82	32,47	14,03

and faster approach but also yields elevated Fitness scores in comparison to state-of-the-art outcomes (Moradi et al., 2023a). Worth highlighting is that the DL model in (Moradi et al., 2023a) incorporates 193,418 learnable parameters, while the proposed method employs only between 662 and 4,924 learnable parameters, averaging around 1,731 (approximately 0.9%). Out of this total, 491 parameters are allocated to TDM, and the quantity varies for TIM depending on its optimized architecture.

Furthermore, it is essential to emphasize that in (Moradi et al., 2023a), the LOOCV procedure employed 11 units for training and 1 unit for validation. That approach yielded results based on the network's output with the best validation loss, which was also the same as the test panel. In contrast, in the current study, the LOOCV process employs different composite panels for validation and testing, thus achieving a higher degree of generalization and practicality.

The comprehensive outcomes of TIM-TDM across various subsets and PCA-based techniques are depicted in Figs. 8 and 9, illustrating the RMSE and Fitness scores, respectively. The results indicate the mean values (displayed 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 (standard deviation) bars.

Fig. 8 shows the RMSE between the simulated ideal HIs and the constructed HIs by the developed model ($HI^{(2)}$). With the exception of specimen 9, the B.1 PCA-based technique generally yields a lower average RMSE compared to others, notably evident in specimens 2, 4, and 5. Following B.1, A.2 (observed in specimens 5 and 10) and B.2 (seen in specimens 8 and 9) demonstrate better RMSE results on average. Yet, it is important to note that RMSE outcomes vary based first on the test unit and then on the validation unit. For instance, B.1 is consistently the best choice for test specimen 5, regardless of the chosen validation specimen. However, this trend might not hold true for test specimen 6, as its superiority depends on the selection of the validation specimen.

Given that the primary objective is to provide qualified HIs based on prognostic criteria, Fig. 9 holds greater significance than Fig. 8. The outcomes presented in Fig. 9, showcasing Fitness scores, exhibit a higher degree of stability compared to the RMSE results in Fig. 8. Across various units (except for 2, 3, and 8), the B.1 PCA-based technique consistently yields higher Fitness scores, while the A.2 PCA version displays the lowest average Fitness score.

The test units that poses the greatest challenge for acceptable PCA versions is 5, where the discrepancy with the impractical version B.1 is substantial. Unit 8 presents challenges across all PCA versions. Notably, in the exceptional case of unit 3, the B.2 PCA version stands out as the most effective. The key point is that the choice of a suitable validation unit can significantly influence the score. For instance, in test unit 11 with the best validation unit 1, the score becomes highly acceptable, even for the A.2 PCA version.

6.2. Ensemble learner models

In this section, the outcomes of ELMs are presented, covering both SAE and WAE, as discussed in Section 5.2. Various ω_k values are explored for WAE, encompassing ω_k^{MSE}, $\omega_k^{Fitness}$, and $\omega_k^{Fitness\_MSE}$ as defined in Eqs. (18)–(20), respectively. For $\omega_k^{Fitness}$, we consider various sets of coefficients {a,b,c}, including {1,1,1}, {1,1,0}, {1,0,0}, and {1,0.5,0.25}, to assess the impact of diverse prognostic metrics. As for $\omega_k^{Fitness\_MSE}$, the coefficient set {1,1,1} is exclusively taken.

Fig. 10 depicts the distributions of the Fitness score across 110 subsets (resulting from 11 validation-index-based combinations × 10 replications) for each fold. The x-axis represents the fold, or, in other words, the single-stiffener composite panel chosen as the test unit for the model. This figure offers insights into the various versions of PCA and ELMs employed. Additionally, **Appendix** contains figures showcasing all the constructed HIs post-EL for different versions of PCA.

As observed in Fig. 10, the EL step, regardless of its type, consistently enhances the score in comparison to the mean value across all cases. This improvement is particularly notable for the A.2 PCA version, while it has a relatively smaller impact on the B.1 PCA version. Specifically, the B.1 PCA version yields a more consistent distribution of highly qualified HIs with less variance, as indicated by the box plot. However, it is crucial to note that this PCA version is unsuitable for prognostics, as previously discussed.

Certain folds present greater challenges for different PCA versions. For instance, specimens 5, 3, and 1 are more challenging for the A.1 PCA version, while specimens 3, 2, 4, and 1 pose challenges for the A.2 PCA version. The B.1 PCA version encounters difficulties with specimens 8 and 3, while the B.2 PCA version struggles with specimens 2 and 5.

While examining various applied EL techniques, it is notable that WAE-MSE does not necessarily outperform SAE. However, WAE-Fitness and WAE-FitnessMSE consistently yield better outcomes than SAE. Focusing on WAE-Fitness, it is evident that the prognostic metrics can sometimes oppose each other. For instance, taking composite panel 6 for the A.2 PCA version (refer to Figure A.2 in Appendix), considering the metrics coefficient set {1,1,0} achieves the desired Pr, an important aspect. Conversely, when the Tr coefficient increases (in cases of {1,1,1} or {1,0.5,0.25}), Pr decreases. When Pr is achieved, the challenge of extrapolation in RUL prediction can be transformed into an interpolation problem, and as is well known, an extrapolation problem still poses a significant challenge in the field of machine learning.

6.3. Discussion

Table 5 presents a comprehensive analysis of the performance of health indicators constructed using various methods from acoustic emission data for composite structures in the ReMAP dataset. The results indicate that the introduction of ensemble learning consistently improved the Fitness score across all cases, regardless of the PCA type employed. Notably, the A.2 PCA version exhibited the most significant benefits from EL, while the B.1 version showed comparatively less improvement. It should be noted that the performance and improvements are reported considering that the maximum Fitness score is 3, corresponding to a performance of 100%. The introduction of TDM further enhanced the Fitness scores compared to TIM, with improvements of up to 8.93%, 11.05%, 13.94%, and 8.08% for PCA versions A.1, A.2, B.1, and B.2, respectively. Additionally, the application of EL using the WAE-FitnessMSE method resulted in even greater enhancements in Fitness scores when compared to TDM.

Given the impracticality of applying the B.1 version, PCA version A.2, characterized by higher stability and the second-highest score, emerges as the optimal choice within the TIM-TDM framework, demonstrating a robust performance of 90% (±2%). Although the SSLSTM approach using AE data (Moradi et al., 2023a) appears to

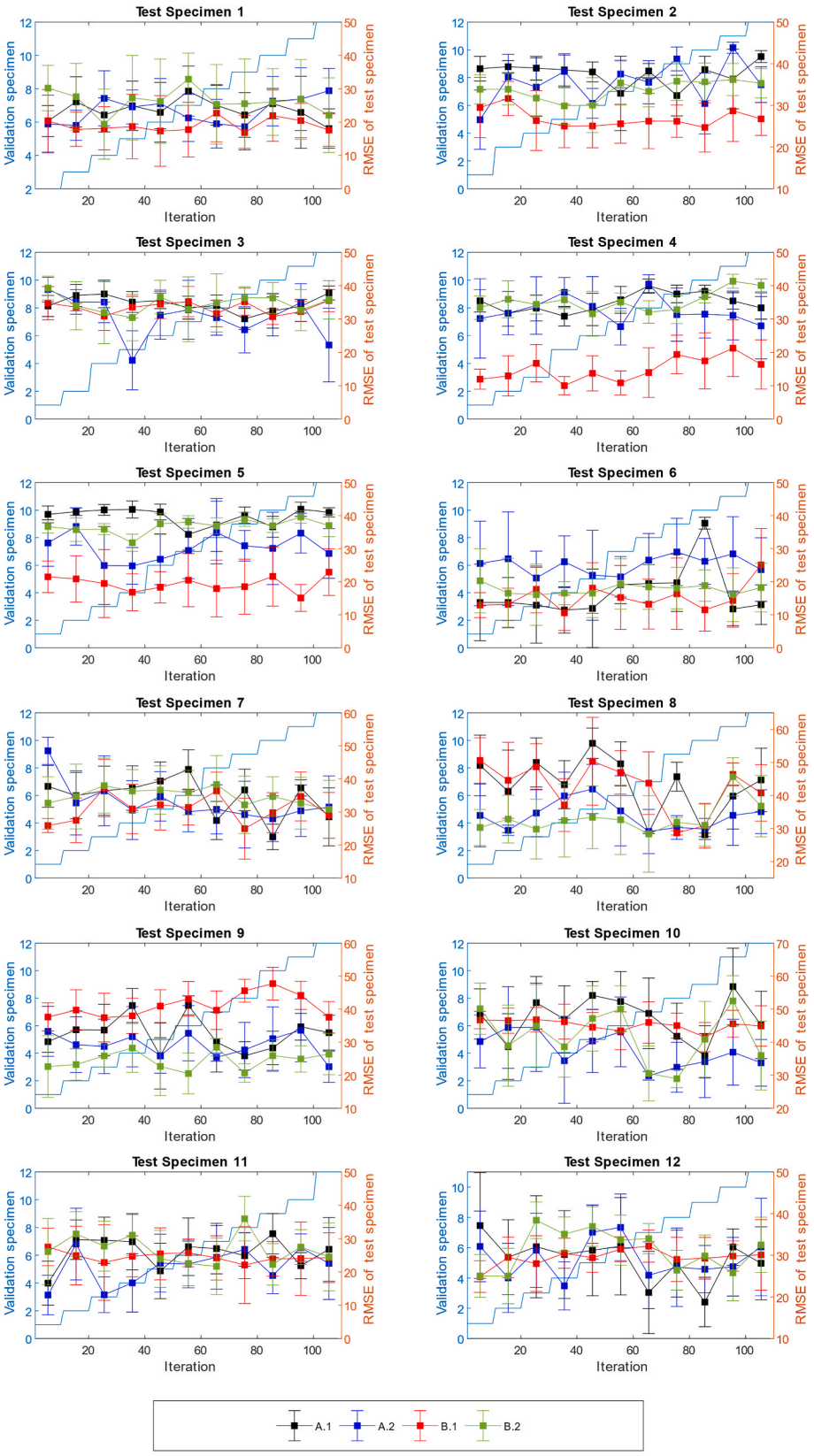

Fig. 8. RMSE between ideal HIs and constructed HIs by the proposed methodology over various subsets (train, validation, and test combinations) and PCA-based techniques (A.1, A.2, B.1, and B.2).

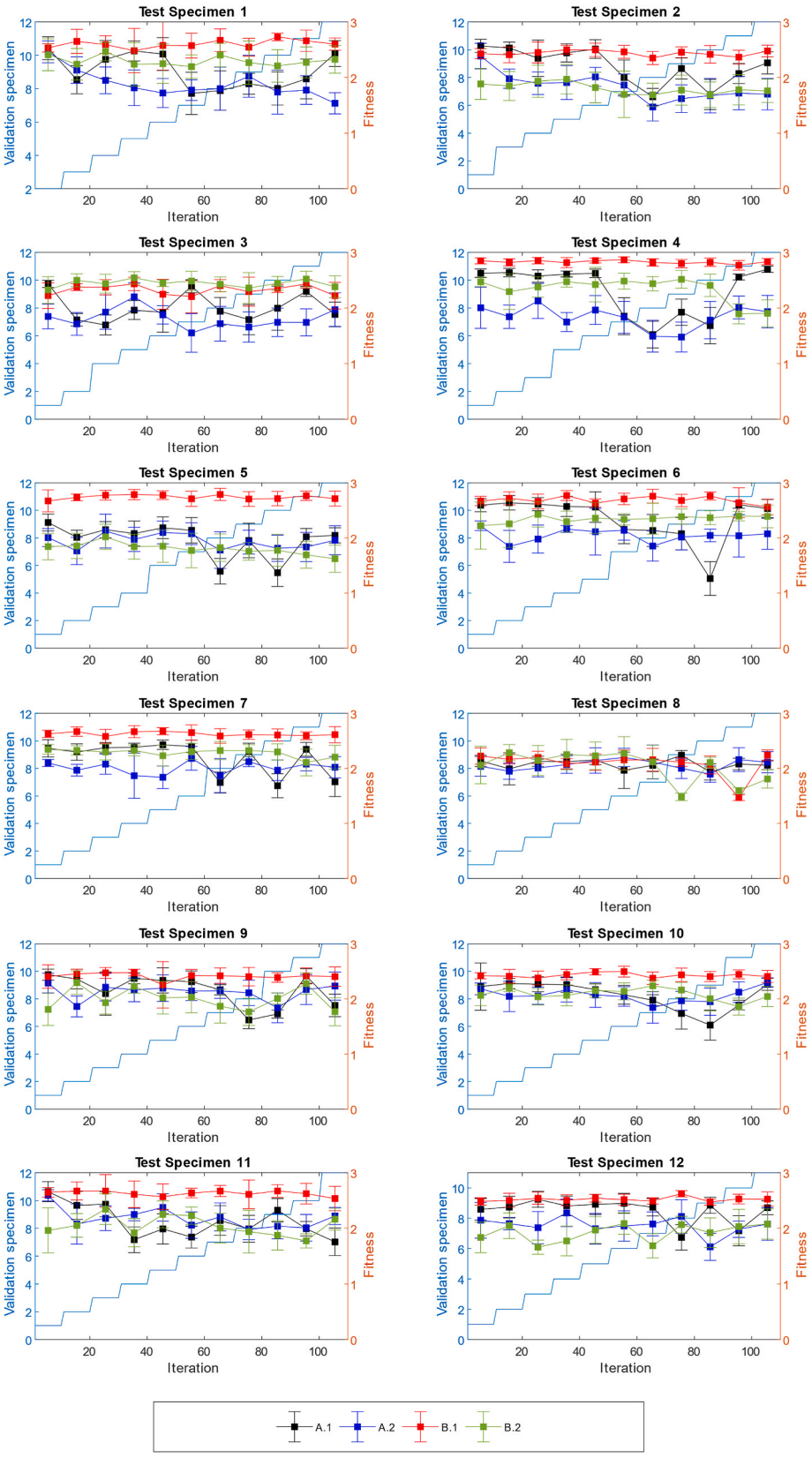

Fig. 9. Fitness for the proposed methodology over various subsets (train, validation, and test combinations) and PCA-based techniques (A.1, A.2, B.1, and B.2).

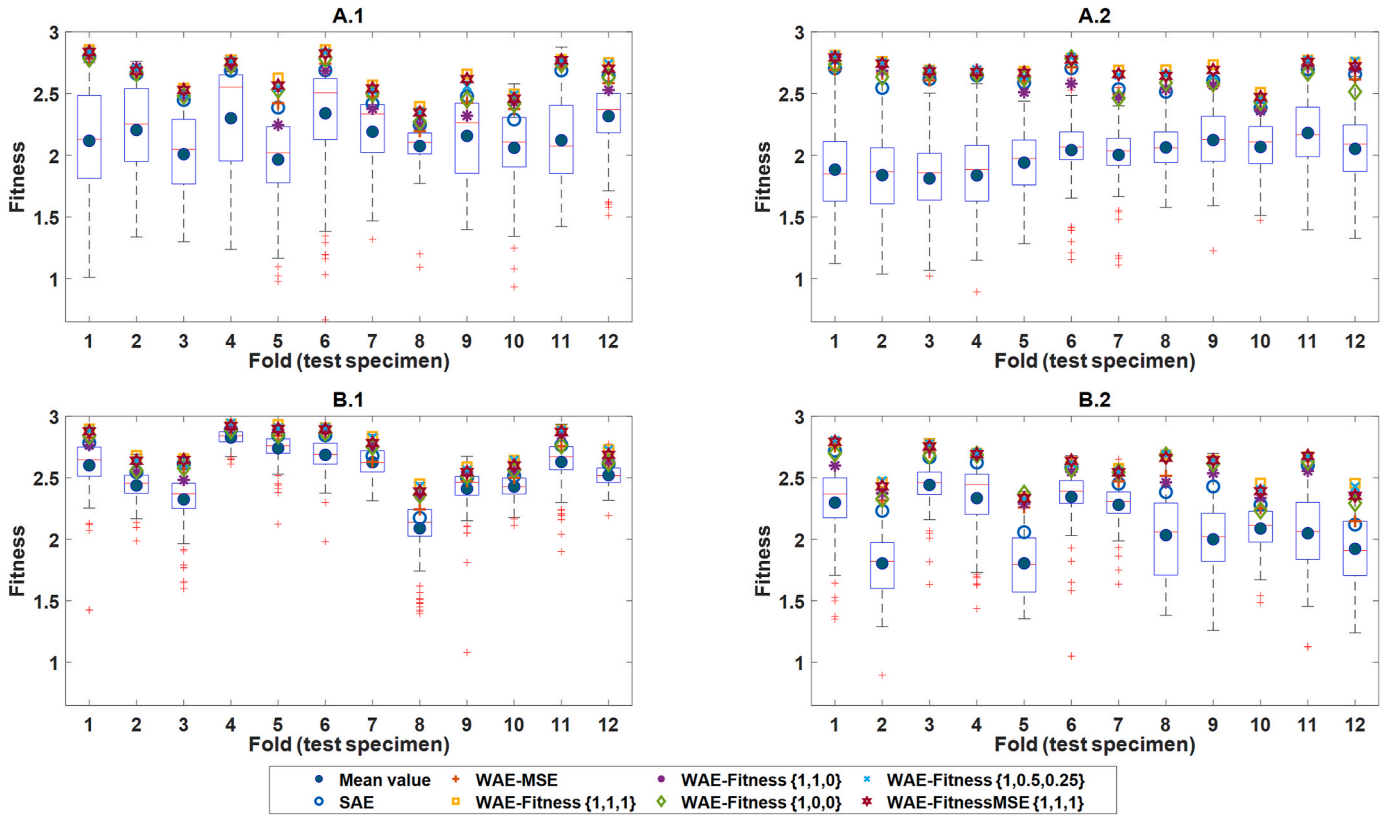

Fig. 10. Distribution of Fitness for the proposed methodology over various subsets (train, validation, and test combinations) and versions of PCA (A.1, A.2, B.1, and B.2). The box plot and the mean value (represented by a dark blue-filled circle) pertain to results before EL, while the other markers depict the scores of the HIs after EL.

Table 5
Performance of health indicators constructed using various methods from acoustic emission data for composite structures in the ReMAP dataset.

PCA type	The average Fitness over 1320 runs for each case					
	TIM	Improvement	TDM	Improvement	WAE-FitnessMSE	SSLSTM (Moradi et al., 2023a) (simple ensemble)
A.1	1.89 (±0.50)	↑8.93%	2.16 (±0.37)	↑17.01%	2.67 (±0.14)	2.79 (±0.14)
A.2	1.66 (±0.39)	↑11.05%	1.99 (±0.31)	↑23.49%	2.69 (±0.07)	
B.1	2.11 (±0.28)	↑13.94%	2.53 (±0.25)	↑6.64%	2.73 (±0.19)	
B.2	1.87 (±0.24)	↑8.08%	2.12 (±0.34)	↑14.99%	2.57 (±0.14)	

provide the best performance at 93%, its applicability in real-world situations where test units are unknown may be limited. This limitation arises from its dependence on test units for validation during the leave-one-out cross-validation phase, which is necessary for stopping training (i.e., determining the optimal number of epochs) and fine-tuning hyperparameters. The current work has addressed this constraint while utilizing significantly fewer learnable parameters (approximately 0.9%). Additionally, with a standard deviation of ±0.07, the performance stability of the suggested framework—the A.2 PCA followed by TIM-TDM and then WAE-FitnessMS—is better than that of SSLSTM, which has a standard deviation of ±0.14, all the while maintaining a high mean performance of 90%.

The main conclusions drawn from this analysis are as follows:

- The B.1 PCA variant consistently yields the highest scores because it leverages information from future observations; this highlights the influence of uncertainties arising from unseen data distributions.
- Results obtained with holdout subsets are encouraging, yet they depend on a judicious choice of training-and-validation units. Certain cross-validation folds—those involving the most challenging panels—prove more difficult. Ensemble learning mitigates this sensitivity by blending models trained on different subsets, thereby reducing the intrinsic randomness of the TIM–TDM pair.
- Ensemble learning improves performance in every case. Among the strategies tested, WAE-Fitness and WAE-FitnessMSE outperform SAE, whereas WAE-MSE does not. In WAE-Fitness, prognostic metrics sometimes conflict.
- PCA variant A.2 offers the best balance of stability and accuracy, delivering a robust performance of 90% (±2%) with the fewest practical constraints.
- Adding TDM further boosts Fitness relative to TIM alone—by 8.93%, 11.05%, 13.94% and 8.08% for PCA variants A.1, A.2, B.1 and B.2, respectively.

These findings underscore the value of integrating advanced ML techniques—including adaptive standardization, dimensionality reduction, and ensemble fusion—for reliable SHM under uncertain conditions, while maintaining dependability in real-world applications.

7. Conclusions

This study introduced a two-stage, intrinsically semi-supervised machine learning framework for constructing health indicators from

acoustic emission data recorded on impact-damaged, stiffened composite panels subjected to uncertainties. Adaptive, per-unit standardization was combined with a variant of PCA to reduce the original 201-dimensional feature space while preserving high variance and avoiding data leakage. A time-independent neural regressor, optimized by physics-guided Bayesian search, generated a first-level health indicator that was subsequently refined by a time-dependent LSTM. Validation on twelve panels, whose fatigue lives ranged from 4.8×10^4 to 7.6×10^5 cycles, produced an average prognostic fitness of 90% $\pm$ 2% under leave-one-out cross-validation with ensemble fusion—an improvement of up to 14% over the single-model baseline.

Four principal contributions can be distilled from this work. First, the paper formalizes an adaptive standardization scheme to mitigate unit-to-unit bias without using future data. Second, it conducts a systematic assessment of four PCA variants and identifies a leakage-free, computationally lightweight configuration suitable for real-time, on-board dimensionality reduction. Third, by partitioning HI construction into a time-independent and a time-dependent stage, the framework exploits the large pool of spatial samples before refining temporal dynamics, markedly improving data efficiency. Fourth, we embed physics-guided Bayesian optimization whose dual loss combines (i) an explicit term that enforces monotonicity, prognosability and trendability and (ii) an implicit term that minimizes deviation from simulated proxy labels, enabling effective learning even when run-to-failure labels are scarce. Together, these elements yield health indicators that are simultaneously monotonic, prognosable and trendable in real time.

Notwithstanding these advances, the framework has certain limitations. The experimental campaign was restricted to specimens sharing similar laminate layups and boundary conditions; investigating a broader set of composite layups, geometries, environmental exposures, and loading spectra would be necessary to establish wider generality. In its current form the method offers a favourable balance between accuracy and computational cost: it tolerates incomplete labels, adapts online to non-stationary signals, and exploits both spatial richness and temporal correlation. Its principal drawbacks are a degree of sensitivity to acoustic emission noise and a dependence on uninterrupted monitoring—if segments of the acoustic emission record are lost, performance cannot be guaranteed.

Future research will extend the dimensionality reduction stage to nonlinear manifold learners, explore the construction of history-independent health indicators from acoustic emission data, integrate additional sensing modalities (e.g., guided waves or distributed strain), and investigate federated learning schemes that allow fleet-wide model adaptation while preserving data privacy.

CRediT authorship contribution statement

Morteza Moradi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Juan Chiachío:** Writing – review & editing, Supervision. **Dimitrios Zarouchas:** Writing – review & editing, Supervision, Conceptualization.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 859957 "ENHAnCE, European training Network in intelligent prognostics and Health mAnagement in Composite structurEs".

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.engappai.2025.112767.

Data availability

The data is publicly available.

References

Ahmadivala, M., 2020. Towards optimal maintenance planning of existing structures based on time-dependent reliability analysis. PhD Thesis. Université Clermont Auvergne.

Ameri, B., Moradi, M., Mohammadi, B., Salimi-Majd, D., 2020. Investigation of nonlinear post-buckling delamination in curved laminated composite panels via cohesive zone model. Thin-Walled Struct. 154, 106797. https://doi.org/10.1016/j.tws.2020.106797.

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Stat. Surv. 4 (none). https://doi.org/10.1214/09-SS054.

Baptista, M.L., Goebel, K., Henriques, E.M.P., 2022. Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. Artif. Intell. 306, 103667. https://doi.org/10.1016/j.artint.2022.103667.

Barabadi, A., Barabady, J., Markeset, T., 2011. Maintainability analysis considering time-dependent and time-independent covariates. Reliab. Eng. Syst. Saf. 96 (1), 210–217. https://doi.org/10.1016/j.ress.2010.08.007.

Beaumont, P.W.R., 2020. The structural integrity of composite materials and long-life implementation of composite structures. Appl. Compos. Mater. 27 (5), 449–478. https://doi.org/10.1007/S10443-020-09822-6/FIGURES/13.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140. https://doi.org/10.1007/BF00058655/METRICS.

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

Chen, T., Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.

Coble, J.B., 2010. Merging Data Sources to Predict Remaining Useful Life–an Automated Method to Identify Prognostic Parameters'. University of Tennessee.

de Moura, E.P., Souto, C.R., Silva, A.A., Irmão, M.A.S., 2011. Evaluation of principal component analysis and neural network performance for bearing fault diagnosis from vibration signal processed by RS and DF analyses. Mech. Syst. Signal Process. 25 (5), 1765–1772. https://doi.org/10.1016/j.ymssp.2010.11.021.

Dunteman, G., 1989. *Principal Components Analysis*. 2455 Teller Road. SAGE Publications, Inc., Newbury Park California 91320 United States of America https://doi.org/10.4135/9781412985475.

El-Thalji, I., Jantunen, E., 2015. A summary of fault modelling and predictive health monitoring of rolling element bearings. Mech. Syst. Signal Process. 60 (61), 252–272. https://doi.org/10.1016/j.ymssp.2015.02.008.

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

Farrar, C.R., Worden, K., 2007. An introduction to structural health monitoring. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 365 (1851), 303–315. https://doi.org/10.1098/rsta.2006.1928.

Ferreira, C., Gonçalves, G., 2022. Remaining Useful Life prediction and challenges: a literature review on the use of machine learning methods. J. Manuf. Syst. 63, 550–562. https://doi.org/10.1016/j.jmsy.2022.05.010.

Foresee, F.D., Hagan, M.T., 1997, June. Gauss-Newton approximation to Bayesian learning. In: Proceedings of international conference on neural networks, (ICNN'97) (Vol. 3. IEEE, pp. 1930–1935.

Freund, Y., Schapire, R.E., 1995. A desicion-theoretic Generalization of on-line Learning and an Application to Boosting, pp. 23–37. https://doi.org/10.1007/3-540-59119-2_166.

Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). The annals of statistics 28 (2), 337–407. https://doi.org/10.1214/aos/1016218223 doi: 10.1214/AOS/1016218223.

Fu, S., Avdelidis, N.P., 2023. Prognostic and health management of critical aircraft systems and components: an overview. Sensors 23 (19), 8124. https://doi.org/10.3390/s23198124.

Galanopoulos, G., Milanoski, D., Broer, A.A.R., Zarouchas, D., Loutas, T., 2021a. Health indicators for diagnostics and prognostics of composite aerospace structures. In: 2021 IEEE 8th International Workshop on Metrology for Aerospace (MetroAeroSpace). IEEE, pp. 541–546. https://doi.org/10.1109/MetroAeroSpace51421.2021.9511759.

Galanopoulos, G., Milanoski, D., Broer, A., Zarouchas, D., Loutas, T., 2021b. Health monitoring of aerospace structures utilizing novel health indicators extracted from complex strain and acoustic emission data. Sensors 21 (17), 5701. https://doi.org/10.3390/S21175701, 2021, *Vol.* 21, *Page* 5701.

Galanopoulos, G., Eleftheroglou, N., Milanoski, D., Broer, A., Zarouchas, D., Loutas, T., 2023. A novel strain-based health indicator for the remaining useful life estimation

of degrading composite structures. Compos. Struct. 306, 116579. https://doi.org/10.1016/J.COMPSTRUCT.2022.116579.
Guo, L., Li, N., Jia, F., Lei, Y., Lin, J., 2017. A recurrent neural network based health indicator for remaining useful life prediction of bearings. Neurocomputing 240, 98–109. https://doi.org/10.1016/j.neucom.2017.02.045.
Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the marquardt algorithm. IEEE Trans Neural Netw 5 (6), 989–993. https://doi.org/10.1109/72.329697.
Hagan, M.T., Demuth, H.B., Beale, M., 1997. Neural Network Design. PWS Publishing Co., USA.
Hartley, R., Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Cambridge, UNITED KINGDOM. Cambridge University Press [Online]. Available: http://ebookcentral.proquest.com/lib/delft/detail.action?docID=256634.
Hashem, S., 1997. Optimal linear combinations of neural networks. Neural Netw. 10 (4), 599–614. https://doi.org/10.1016/S0893-6080(96)00098-6.
Hastie, T., Friedman, J., Tibshirani, R., 2001. The Elements of Statistical Learning. Springer New York, New York, NY. https://doi.org/10.1007/978-0-387-21606-5.
Haykin, S., 1998. Neural Networks: a Comprehensive Foundation. Prentice Hall PTR. https://doi.org/10.1016/0967-0661(95)90080-2.
He, Y., et al., 2021. An overview of acoustic emission inspection and monitoring technology in the key components of renewable energy systems. Mech. Syst. Signal Process. 148, 107146. https://doi.org/10.1016/J.YMSSP.2020.107146.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.
Huber, P.J., 1992. Robust Estimation of a Location Parameter, pp. 492–518. https://doi.org/10.1007/978-1-4612-4380-9_35.
Kamranfar, P., 2023. Machine Learning Enabled Health Monitoring and Diagnosis of Engineering Systems. George Mason University.
Kingma, D.P., Ba, J.L., 2015. Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.
Kralovec, C., Schagerl, M., 2020. Review of structural health monitoring methods regarding a multi-sensor approach for damage assessment of metal and composite structures. Sensors 20 (3), 826. https://doi.org/10.3390/s20030826.
Lei, Y., 2016. Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery. https://doi.org/10.1016/c2016-0-00367-4.
Liu, Y., Mohanty, S., Chattopadhyay, A., 2009. In: Lindner, D.K. (Ed.), A Gaussian Process Based Prognostics Framework for Composite Structures, p. 72860J. https://doi.org/10.1117/12.815889.
MacKay, D.J.C., 1992. Bayesian interpolation. Neural Comput. 4 (3), 415–447. https://doi.org/10.1162/neco.1992.4.3.415.
Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters. J. Soc. Ind. Appl. Math. 11 (2), 431–441. https://doi.org/10.1137/0111030.
Marsland, S., 2011. Machine Learning: an Algorithmic Perspective, first ed. Chapman and Hall/CRC. https://doi.org/10.1201/9781420067194.
Matsumoto, M., Nishimura, T., 1998. Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model Comput. Simulat 8 (1), 3–30. https://doi.org/10.1145/272991.272995.
Milanoski, D.P., Loutas, T.H., 2021. Strain-based health indicators for the structural health monitoring of stiffened composite panels. J. Intell. Mater. Syst. Struct. 32 (3), 255–266. https://doi.org/10.1177/1045389X20924822.
Moradi, M., 2024. Designing Health Indicators for Aerospace Structures by Intelligent Information Fusion. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:7ac03701-b97a-427d-990c-e6c696d1254b.
Moradi, M., Komninos, P., Benedictus, R., Zarouchas, D., 2022a. Interpretable neural network with limited weights for constructing simple and explainable HI using SHM data. In: Annual Conference of the PHM Society. https://doi.org/10.36001/phmconf.2022.v14i1.3185.
Moradi, M., Broer, A., Zarouchas, D., 2022b. Acoustic emission dataset of single-stiffener composite panels subjected to impact and run-to-failure fatigue loading. Mendeley Data. https://doi.org/10.17632/YS8R8M7BX2.2.
Moradi, M., Broer, A., Chiachío, J., Benedictus, R., Loutas, T.H., Zarouchas, D., 2023a. Intelligent health indicator construction for prognostics of composite structures utilizing a semi-supervised deep neural network and SHM data. Eng. Appl. Artif. Intell. 117, 105502. https://doi.org/10.1016/J.ENGAPPAI.2022.105502.
Moradi, M., Broer, A., Chiachío, J., Benedictus, R., Zarouchas, D., 2023b. Intelligent health indicators based on semi-supervised learning utilizing acoustic emission data. Lecture Notes in Civil Engineering 270, 419–428. https://doi.org/10.1007/978-3-031-07322-9_43. LNCE.
Moradi, M., Komninos, P., Zarouchas, D., 2025. Constructing explainable health indicators for aircraft engines by developing an interpretable neural network with discretized weights. Appl. Intell. 55 (2), 143.
Moradi, M., Gul, F.C., Zarouchas, D., 2024b. A novel machine learning model to design historical-independent health indicators for composite structures. Compos. B Eng. 275, 111328. https://doi.org/10.1016/j.compositesb.2024.111328.
Nguyen, K.T.P., Medjaher, K., 2021. An automated health indicator construction methodology for prognostics based on multi-criteria optimization. ISA Trans. 113, 81–96. https://doi.org/10.1016/j.isatra.2020.03.017.
Nuhic, A., Terzimehic, T., Soczka-Guth, T., Buchholz, M., Dietmayer, K., 2013. Health diagnosis and remaining useful life prognostics of lithium-ion batteries using data-driven methods. J. Power Sources 239, 680–688. https://doi.org/10.1016/j.jpowsour.2012.11.146.
P, L.E.O., Van Der Maaten, H.J.V.D.H., 2009. Dimensionality reduction: a comparative review. Journal of machine learning research 10, 66–71.
Opitz, D., Maclin, R., 1999. Popular ensemble methods: an empirical Study. J. Artif. Intell. Res. 11, 169–198. https://doi.org/10.1613/JAIR.614.
Peng, T., Liu, Y., Saxena, A., Goebel, K., 2015. In-situ fatigue life prognosis for composite laminates based on stiffness degradation. Compos. Struct. 132, 155–165. https://doi.org/10.1016/J.COMPSTRUCT.2015.05.006.
Polikar, R., 2006. Ensemble based systems in decision making. IEEE Circ. Syst. Mag. 6 (3), 21–44. https://doi.org/10.1109/MCAS.2006.1688199.
Riedmiller, M., Braun, H., 1993, March. A direct adaptive method for faster backpropagation learning: The RPROP algorithm. In: IEEE international conference on neural networks. IEEE, pp. 586–591.
Saidi, L., Ben Ali, J., Bechhoefer, E., Benbouzid, M., 2017. Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR. Appl. Acoust. 120, 1–8. https://doi.org/10.1016/j.apacoust.2017.01.005.
Saxena, A., et al., 2008. Metrics for evaluating performance of prognostic techniques. In: 2008 International Conference on Prognostics and Health Management. IEEE, pp. 1–17. https://doi.org/10.1109/PHM.2008.4711436.
Saxena, A., Celaya, J., Saha, B., Saha, S., Goebel, K., 2021. Metrics for offline evaluation of prognostic performance. Int J Progn Health Manag 1 (1). https://doi.org/10.36001/ijphm.2010.v1i1.1336.
Sbarufatti, C., Manes, A., Giglio, M., 2013. Performance optimization of a diagnostic system based upon a simulated strain field for fatigue damage characterization. Mech. Syst. Signal Process. 40 (2), 667–690. https://doi.org/10.1016/j.ymssp.2013.06.003.
Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. In: Pereira, F., Burges, C.J., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems. Curran Associates, Inc. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf
Suresh Kumar, C., Fotouhi, M., Saeedifar, M., Arumugam, V., 2019. Acoustic emission based investigation on the effect of temperature and hybridization on drop weight impact and post-impact residual strength of hemp and basalt fibres reinforced polymer composite laminates. Compos. B Eng. 173, 106962. https://doi.org/10.1016/j.compositesb.2019.106962.
Valkonen, A.J., 2023. Exploring Structural Health Monitoring Value of Information Based on Remaining Useful Life Extension Potential'. Princeton University.
van Engelen, J.E., Hoos, H.H., 2020. A survey on semi-supervised learning. Mach. Learn. 109 (2), 373–440. https://doi.org/10.1007/s10994-019-05855-6.
Wolpert, D.H., 1992. Stacked generalization. Neural Netw. 5 (2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1.
Wu, F., Yao, W., 2010. A fatigue damage model of composite materials. Int. J. Fatig. 32 (1), 134–138. https://doi.org/10.1016/j.ijfatigue.2009.02.027.
Yu, H., Li, H., 2022. Pump remaining useful life prediction based on multi-source fusion and monotonicity-constrained particle filtering. Mech. Syst. Signal Process. 170, 108851. https://doi.org/10.1016/J.YMSSP.2022.108851.
Yuan, F.-G., Zargar, S.A., Chen, Q., Wang, S., 2020. Machine learning for structural health monitoring: challenges and opportunities. In: Zonta, D., Sohn, H., Huang, H. (Eds.), Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2020. SPIE, p. 2. https://doi.org/10.1117/12.2561610.
Yue, N., Broer, A., Briand, W., Rébillat, M., Loutas, T., Zarouchas, D., 2022. Assessing stiffness degradation of stiffened composite panels in post-buckling compression-compression fatigue using guided waves. Compos. Struct. 293, 115751. https://doi.org/10.1016/J.COMPSTRUCT.2022.115751.
Zarouchas, D., Broer, A., Galanopoulos, G., Briand, W., Benedictus, R., Loutas, T., 2021. Compression Compression Fatigue Tests on Single Stiffener Aerospace Structures. https://doi.org/10.34894/QNURER.