Intelligent health indicator construction for prognostics of composite structures utilizing a semi-supervised deep neural network and SHM data

Moradi, Morteza; Broer, Agnes; Chiachío, Juan; Benedictus, Rinze; Loutas, Theodoros H.; Zarouchas, Dimitrios

DOI
10.1016/j.engappai.2022.105502

Publication date
2023

Document Version
Final published version

Published in
Engineering Applications of Artificial Intelligence

Citation (APA)
Moradi, M., Broer, A., Chiachío, J., Benedictus, R., Loutas, T. H., & Zarouchas, D. (2023). Intelligent health indicator construction for prognostics of composite structures utilizing a semi-supervised deep neural network and SHM data. *Engineering Applications of Artificial Intelligence*, *117*, Article 105502. https://doi.org/10.1016/j.engappai.2022.105502

Engineering Applications of Artificial Intelligence 117 (2023) 105502

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Intelligent health indicator construction for prognostics of composite structures utilizing a semi-supervised deep neural network and SHM data

Morteza Moradi [a,b,*], Agnes Broer [a,b], Juan Chiachío [c], Rinze Benedictus [a], Theodoros H. Loutas [d], Dimitrios Zarouchas [a,b]

[a] Structural Integrity & Composites Group, Aerospace Engineering Faculty, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, P.O. Box 5058, 2600 GB Delft, The Netherlands

[b] 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

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

[d] Laboratory of Applied Mechanics and Vibrations, Department of Mechanical and Aeronautics Engineering, University of Patras, University Campus, 265 04 Rio Patras, Greece

ARTICLE INFO

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

ABSTRACT

A health indicator (HI) is a valuable index demonstrating the health level of an engineering system or structure, which is a direct intermediate connection between raw signals collected by structural health monitoring (SHM) methods and prognostic models for remaining useful life estimation. An appropriate HI should conform to prognostic criteria, i.e., monotonicity, trendability, and prognosability, that are commonly utilized to measure the HI's quality. However, constructing such a HI is challenging, particularly for composite structures due to their vulnerability to complex damage scenarios. Data-driven models and deep learning are powerful mathematical tools that can be employed to achieve this purpose. Yet the availability of a large dataset with labels plays a crucial role in these fields, and the data collected by SHM methods can only be labeled after the structure fails. In this respect, semi-supervised learning can incorporate unlabeled data monitored from structures that have not yet failed. In the present work, a semi-supervised deep neural network is proposed to construct HI by SHM data fusion. For the first time, the prognostic criteria are used as targets of the network rather than employing them only as a measurement tool of HI's quality. In this regard, the acoustic emission method was used to monitor composite panels during fatigue loading, and extracted features were used to construct an intelligent HI. Finally, the proposed roadmap is evaluated by the holdout method, which shows a 77.3% improvement in the HI's quality, and the leave-one-out cross-validation method, which indicates the generalized model has at least an 81.77% score on the prognostic criteria. This study demonstrates that even when the true HI labels are unknown but the qualified HI pattern (according to the prognostic criteria) can be recognized, a model can still be built that provides HIs aligning with the desired degradation behavior.

1. Introduction

Prognostic and health management (PHM) is a natural extension of structural health monitoring (SHM) in the sense that the predictions of remaining useful life (RUL) are timely updated using data from sensors. To be able to predict RUL, a Health Indicator (HI) suitable to enter into a prognostic model is needed (Galanopoulos et al., 2021b). HI is a special feature extracted from SHM data that shows the health (or damage) status of the structure or system under monitoring. If the HI exceeds a specific threshold, the SHM system should alarm in order to proceed with the relevant guidelines such as system shutdown, maintenance, or replacement. As a result, HI provides a connection and is a prominent feature between raw signals and the prognostic model, and it has a direct relationship with RUL. In contrast to RUL, which is commonly assumed to be a linear or piece-wise linear degradation model (Al-Dulaimi et al., 2019), HIs are nonlinear due to the nonlinear nature of damage propagation and accumulation, which makes it useful for analyzing and connecting to the mechanical behavior of the structure.

Abbreviations: PHM, Prognostic and health management; SHM, Structural health monitoring; IHI, Intelligent health indicator; SSDNN, Semi-supervised deep neural network

* Corresponding author at: Structural Integrity & Composites Group, Aerospace Engineering Faculty, Delft University of Technology, Kluyverweg 1, 2629 HS Delft, P.O. Box 5058, 2600 GB Delft, The Netherlands.

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

https://doi.org/10.1016/j.engappai.2022.105502
Received 24 June 2022; Received in revised form 20 September 2022; Accepted 30 September 2022
Available online xxxx
0952-1976/

On the other hand, HIs are not solely sensitive to damage, but they might be sensitive to abnormalities in environmental conditions and operations. The suitability of HI, which is made of features extracted from different SHM systems, can be determined by criteria such as Monotonicity (Mo — measuring the general increasing or decreasing pattern of a variable over time), Prognosability (Pr — measuring the distribution of a variable's final value), and Trendability (Tr — measuring the similarity between trajectories of variables) (Coble and Hines, 2009; Coble, 2010; Lei, 2016; Saidi et al., 2017; Eleftheroglou, 2020). The main goal and contribution of this paper is the development of a methodology to construct a HI satisfying the above-mentioned prognostic criteria suited for use in prognostic models, which deals with Feature Extraction (FE), Feature Selection (FS), and particularly, Feature Fusion (FF).

The common procedure to design a HI that can be employed to predict the RUL of an engineering system is to select the best features (after FE) in accordance with prognostic criteria as the HI or the main constitutive components of the HI. In this scenario, some features will be overlooked since they do not meet the criteria, while they might be useful since their fusion may comply with the intended specifications. To overcome this shortcoming, the prognostic criteria can play a supervising role in the construction process of HI rather than being only a measurement tool of HI's quality. For example, a predefined function with a set of polynomial components (Eleftheroglou, 2020) can be considered to fuse features, in which the coefficients of the polynomial components are unknown and have to be determined. In this regard, combining the prognostic criteria into a "fitness" function could be considered as an objective function for an optimization problem (Eleftheroglou et al., 2018). In this approach, although a polynomial series might construct a proper HI function, the other components, e.g. logarithmic and exponential ones, might construct a function with more monotonic, trendable, and prognosable behavior. In fact, the fusion function is limited to only polynomial operators in this approach, whereas other mathematical operators and combinations may produce better HI. Also, a significant and critical point in the predefined function scenario to fuse features that should be seriously noted is the computation time. Since the extracted features might be more than hundreds (like the current work), this model is very time-consuming. Thus, a fusion model based on an Artificial Neural Network (ANN), especially Deep Neural Network (DNN), rather than the predefined functions is proposed in the current study.

ANN and DNN are applicable in the field of PHM (Fink et al., 2020; Khan and Yairi, 2018; Zhao et al., 2019) and are powerful mathematical methods for approaching this problem; however, implementation of the criteria directly through an ANN is problematic since backpropagation requires the criteria's derivatives, which are not easy to calculate. Furthermore, because the values of HIs are unknown to be used as targets for ANN, even after end-of-life (EOL), the problem cannot be solved by a supervised learning model. Therefore, the lack of an approach to tackle this challenge is obvious. In fact, the HI construction utilizing FF lies between supervised and unsupervised learning, enabling Semi-Supervised Learning (SSL) to be a feasible option. SSL makes it feasible for enormous volumes of unlabeled data to be exploited in conjunction with smaller labeled data sets (Van Engelen and Hoos, 2020) where it is assumed (in the current research) that at least the HI value is known at EOL. However, the key problem is figuring out how to incorporate Mo, Tr, and Pr metrics into an SSL model, which will be addressed in the present paper.

SSL models have recently attracted interest concerning RUL prediction. He et al. (2022) extended a semi-supervised Generative Adversarial Network (GAN) regression model to take both failure and suspension histories into account in this respect. Also, a robustness evaluation approach is applied to consider the prognostic model's uncertainty induced by the shortage of failure data. Two case studies are used to verify the provided approach's accuracy and reliability. Ellefsen et al. (2019) use a semi-supervised setup to analyze the effects of unsupervised pre-training on RUL estimations, using a Genetic Algorithm (GA) approach to optimize the hyperparameters. According to their observations, unsupervised pre-training is a viable strategy for RUL prediction under a variety of operational conditions and failure types. Based on the literature, SSL has been a promising method for RUL prediction with a (piece-wise) linear degradation trend, but no research has been performed on the HI construction by fusing features using a SSL platform to obey the prognostic criteria. Since only the prognostic criteria and EOL are available to guide the network, an inductive SSL called *intrinsically semi-supervised* (Van Engelen and Hoos, 2020), which is an extension of existing supervised methods to include unlabeled samples in the objective function, is proposed to construct the HI for the first time in the current research. To optimize hyperparameters, a Bayesian optimization algorithm has been used, considering a holdout validation, and the results have been evaluated by Leave-One-Out Cross-Validation (LOOCV) in the perspective of generalization.

To validate the proposed method, acoustic emission (AE) data has been collected from twelve single-stiffener composite panels during compression–compression (C-C) fatigue loading, which are related to the H2020 ReMAP.[1] project. The application to such composite case study poses additional challenges in the construction of HIs for prognostic purposes. Due to the inhomogeneity and design of the composite panels, their damage degradation process is complex. As such, the relation between the collected AE data and the damage degradation is not obvious. Furthermore, the extensive efforts required in composite testing results in few samples and a small available training dataset. Consequently, its application to composite panels makes the objective of HI design and RUL prediction based on AE data challenging. Six AE variables including amplitude, rise time, duration, energy, counts, and RMS were extracted as main data and referred to as "signal" hereinafter. Then, the FE process considering time and frequency domains was carried out, and the resulting features have been considered as the inputs of a Semi-Supervised Deep Neural Network (SSDNN). Although AE as a powerful SHM method was selected as a validation case for the proposed method, the same procedure can be generalized for other SHM methods.

The contributions of this work include:

1. In order to choose the most effective function to generate labels for HI, four functions are assessed based on the Mo, Tr, and Pr.
2. All three above-mentioned prognostic criteria are induced into a SSDNN framework through the simulated labels by the above-chosen function. In this way, Mo, Tr, and Pr implicitly force the neural network to predict the proper HI desirable for prognostic.
3. The Bayesian optimization method is employed to identify the optimal hyperparameters and architectures for the proposed SSDNN. Furthermore, leave-one-out cross-validation is implemented to validate and generalize the approach.
4. Acceptable HIs for composite skin-stiffener panels under C–C fatigue loading using acoustic emission data are provided, which for the first time conform to all three Mo, Tr, and Pr.
5. The findings demonstrate that, in terms of HI's overall (quality) fitness score, the proposed framework outperforms the recently published research on the ReMAP dataset. Additionally, when compared to the features used as HIs introduced by other research, the proposed model provides better HIs according to the prognostic criteria.

The remainder of the paper is organized as follows. It starts with Section 2 which presents and discusses the *Background on health indicators* for HI extraction/construction. Then, the *Experimental Campaigns* will be introduced in Section 3. In Section 4, *Methods* will be described, including *Data Acquisition*, *Pre-processing*, *Signal Processing (SP)*, *Feature Extraction (FE)*, and *Feature Fusion (FF)*. *Feature Fusion (FF)* as the main contribution of the current research contains *Criteria of Prognostic*

[1] ReMAP: Real-time Condition-based Maintenance for Adaptive Aircraft Maintenance Planning. https://h2020-remap.eu/.

Parameters and *Semi-Supervised Criteria-based Fusion Neural Network*. Afterward, *Results and Discussions* will be explained in Section 5, including *Holdout Validation*, *Leave-One-Out Cross Validation (LOOCV)*, and *Discussion*. Section 6 (*Conclusions*) concludes this paper and gives an outlook on future work.

2. Background on health indicators

In order to visualize data and continuously characterize the health state of the structure over its entire life, HIs can be informative and helpful (Song et al., 2017). The available data-driven RUL prognostic methods, such as artificial intelligence and statistical-based models, can also be effectively integrated with the fusion-based HIs (Lei et al., 2018). Depending upon whether a HI has any physical senses, system performance data is typically divided into two categories: the Physical Health Indicator (PHI) and the Virtual Health Indicator (VHI) (Hu et al., 2012).

Since health management has been given more attention and begun earlier for some particular industrial components, such as batteries and rotating machinery equipment, PHIs are relatively well known and helpful. In other words, the long-term and vital use of such components, as well as their history in the industry, in turn results in providing more datasets as well as more solid physical, analytical backgrounds. For instance, according to physical knowledge about batteries, the capacity data of lithium-ion batteries can be considered as a proper HI (Cadini et al., 2019). Pei et al. (2022) recently said that, "To the best of our knowledge, the capacity is a typical performance indicator to monitor the health status of the battery and determine whether the battery requires replacement; thus, we adopt such indicator for the RUL prediction". For rolling element bearings, Huang et al. (2020) utilized Root Mean Square (RMS) as "a simple and practical HI, which is widely used in bearings residual life prediction". Also, Relative RMS (RRMS) could be more robust to characterize the degradation process of bearings (Zhang et al., 2023). However, to the best of our knowledge, no certain and promising PHI has yet been developed for composite structures, especially for structures subjected to C–C fatigue loading, which is one of the most critical and dangerous conditions for composite structures. This can be due to complex scenarios of progressive damage and different types of damage, which in turn are dependent on the type of fibers, matrix, fabrication, curing, boundary and environmental conditions, loading, etc., which is not the case in isotropic material-made components.

Despite the fact that various articles have introduced PHIs (axial strain) utilizing DIC data (Eleftheroglou et al., 2016), or the size or number of cracks have been considered as PHIs (Li et al., 2022), they are not appropriate for genuine SHM and implementations. Concerning the first subject, methods like DIC are not yet considered SHM (Zarouchas and Eleftheroglou, 2020) because of their numerous drawbacks, including the necessity of painting, particular illumination requirements, problematic calibration, etc. The DIC method is also limited to measuring the strain or the deformation at the surface of the structure in the field of view of the camera, making it more effective for plates due to the lack of access to the rear of the surface. Such methods are usually used to validate other aspects of research, such as finite element modeling (Seon et al., 2015) or other NDT/SHM techniques (Saeedifar et al., 2022). For the latter subject, it may be possible to correctly monitor and take into account the number or size of fractures as a HI in isotropic materials such as aluminum (Li et al., 2022), but this is not easily practical for composite materials and complicated structures. In reality, a variety of impactful micro and macro-damage will emerge, such as cracks and delaminations, which not only emerge randomly throughout the structure (Qian et al., 2013) but are also concealed in various quantities among the various layers of the composite material (Ameri et al., 2020).

In contrast to PHI, VHI cannot be easily interpreted and realized in order to make an understandable connection to the physical implications. However, VHI can be designed and optimized based on the intended purposes, such as prognostics. For example, if a monotonic, trendable, and prognosable behavior can be embedded into an objective function (Eleftheroglou, 2020) which is supposed to be used within a data-driven model, the resulted VHI is suited to the next step, the prognostic model. Linear-based feature extraction and selection methods are suitable enough to provide an acceptable HI for some applications. For instance, Principal Component Analysis (PCA) applied to gear vibration signals was able to extract a proper HI (Qin et al., 2019). This method was also used to successfully construct the HI for the CMAPSS dataset (the turbofan engine degradation dataset), the PHM08 dataset (Prognostics Data Challenge Dataset), and the N-CMAPSS dataset (the new CMAPSS dataset) (Song et al., 2022). However, the PCA method does not generate suitable HIs for the ReMAP dataset. Fig. 1 demonstrates the 1st principal component obtained by the PCA model on the CMAPSS dataset in comparison with the ReMAP dataset. It is clear that, in accordance with the Mo, Tr, and Pr, this method is ineffective for building HI utilizing either AE signals (amplitude, rise time, duration, energy, counts, and RMS) or AE features (the features that have been extracted and will be discussed in the current study).

The commonly used nonlinear methods for fusion include kernel methods and DNN. For kernel methods, the kernel function is used to measure the similarity between collected measurements, which is time-consuming and memory-consuming (Wen et al., 2021). On the other hand, DNN is a promising solution if enough labeled data is available. For some cases, like the current work, not only is enough data unavailable, but they are also unlabeled. Thus, by taking advantage of FE on the input side (which leads to providing the informative and labelable inputs) and inducing the intended desirable behavior on the output side of the model (which leads to simulating labels), we try to approach the problem.

It is important to note that an identified HI and its corresponding preprocessing technique are closely relevant to the engineering system/structure under monitoring (Wen et al., 2021), as well as the type of SHM system and sensor that produce signals, and as a result, they cannot be applied to other components directly. This marks the first challenge in comparing the current work, which focuses on a new experiment, to earlier research. In several works, only Mo was reported for the utilized HIs (Eleftheroglou et al., 2018; Zarouchas and Eleftheroglou, 2020; Eleftheroglou and Loutas, 2016), while only one work (Yue et al., 2022) provided all the criteria. The prognostic criteria have therefore infrequently been quantified for the reported HIs, especially for composite structures, which adds another challenge when comparing the current study with existing work. Yue et al. (2022) employed the Guided-Wave (GW) monitoring method in the ReMAP project to predict the stiffness of composite panels, which is a mechanical characteristic of the structure, and then they reported the prognostic criteria for predictions and stiffness. In contrast to the current study, they only considered five samples out of twelve, and it should be highlighted that the more samples used, the lower the fitness score (quality) of HIs. One significant and critical point in the HI construction or RUL prediction frameworks, which is directly related to the Pr, is that input/features data must not be normalized in accordance with the mean and standard deviation of the entire dataset (training and test) (Peng et al., 2019; Wang et al., 2022), as test data are unavailable in reality in the upcoming timeframes. A max–min normalization technique using the full dataset has similar or even more concerns (Jiang et al., 2021; Huang et al., 2019). Therefore, only the mean and standard deviation of the training dataset are used to perform zero-mean normalization in the current study. Several features extracted from AE data, which were considered as HI of the composite structure based on the literature (including 1/A (Eleftheroglou et al., 2018; Zarouchas and Eleftheroglou, 2020), energy (Eleftheroglou et al., 2020), and Rise-time/Amplitude ratio (RA) (Galanopoulos et al., 2021a; Loutas et al., 2017) cumulated in the time window), will be compared with the proposed HI. Also, the prognostic criteria for the proposed AE-based HI will be compared with the HIs extracted from guided wave data and mechanical properties (stiffness) of the ReMAP dataset (Yue et al., 2022).

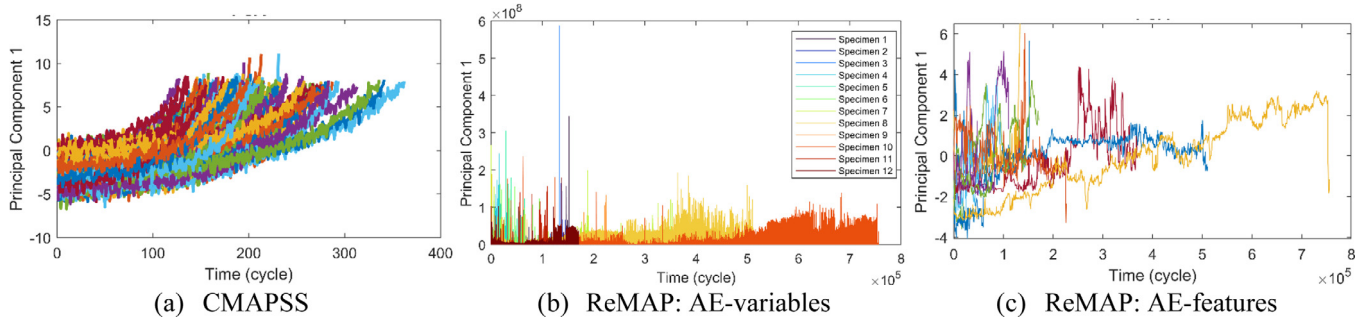

Fig. 1. First principal component calculated using PCA on (a) raw data in the CMAPSS dataset, (b) acoustic emission variables (signals) in the ReMAP dataset, and (c) acoustic emission features in the ReMAP dataset.

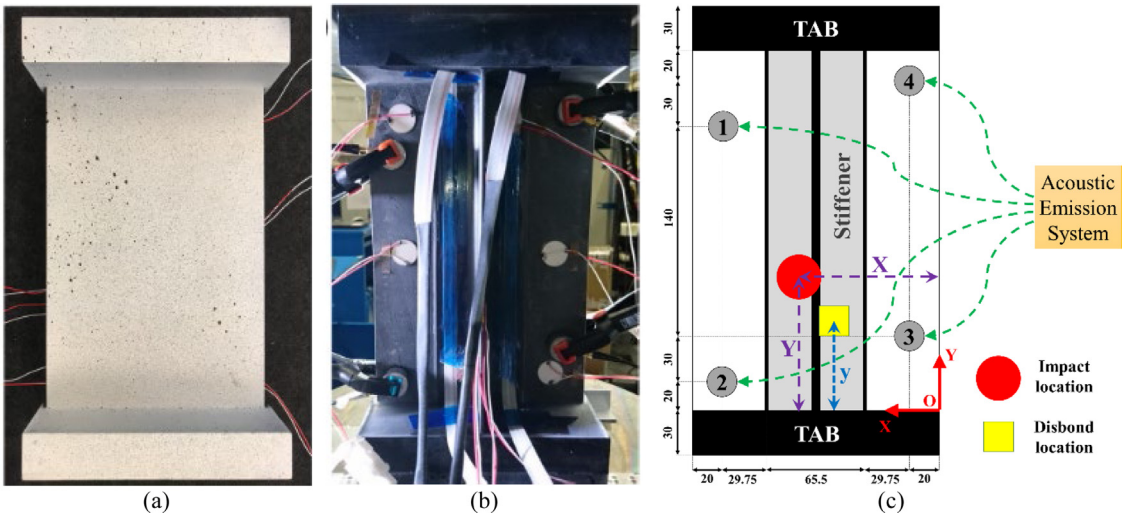

Fig. 2. Single stiffener panel: (a) skin side, (b) stiffener side (Broer et al., 2020), (c) sensor coordinates (dimensions in [mm]), where the four AE sensors shown as gray circles, impact and disbond locations shown as a movable red circle and yellow rectangle for different panels.

3. Experimental campaigns

As part of the H2020 ReMAP project, two test campaigns were held at the Delft Aerospace Structures and Materials Laboratory (DASML) in 2019 and 2020, in which twelve composite skin-stiffener panels were tested under C–C fatigue loading. The panels are made up of a skin panel and a single T-stiffener according to an Embraer design (see Fig. 2(a–b)). The skin and stiffener are all made of IM7/8552 carbon fiber-reinforced epoxy unidirectional prepreg with layups of $[45/-45/0/45/90/-45/0]_S$ and $[45/-45/0/45/-45]_S$, respectively Broer et al. (2020). Two resin blocks for each composite specimen (CS) were also included to ensure that the load was distributed evenly. The dimensions of one panel are shown in Fig. 2(c). In all panels, an impact loading (10 J) has been applied in different locations and times in the stiffener region, and three composite panels also contain disbond defects introduced during manufacturing between the skin panel and the stiffener (more information in Table 1).

The damage growth in the panel was monitored using six different techniques: (1) AE, (2) Distributed Fiber Optic Sensing (DFOS), (3) FBGs (only for the campaign 2019), (4) Lamb Wave Detection System (LWDS) (only for the campaign 2019), (5) Digital Image Correlation (DIC), and (6) Camera. Only data from the AE method is analyzed for the purposes of this research. More information can be explored in Refs. Broer et al. (2020, 2021) and Zarouchas et al. (2021).

The panels were loaded in C–C fatigue loading using an MTS machine with a frequency of 2 Hz and an R-ratio of 10 which means that the fatigue load was set in a compression load range of [6.5, 65.0] kN. Although the R-ratio was intended to hold invariant, the panels experienced a loss in load-bearing capacity. The fatigue load was disrupted at regular intervals of 500 cycles to allow the SHM systems to take measurements. Table 1 summarizes the aforementioned explanations and provides additional details.

4. Methods

Fig. 3 depicts the overall workflow of prognostics, including its main steps. The successive processes from data acquisition to HI are outlined in this section.

4.1. Data acquisition

The AE sensors used are Vallen Systeme GmbH VS900-M broadband sensors with a frequency range of 100–900 kHz. The AE hits were recorded using an AMSY-6 Vallen acquisition system. Moreover, Vallen preamplifiers with a gain of 34 dB were used. Four AE sensors were clamped in various positions on the skin of the panels to create a parallelogram, enabling to localize damage and to obtain a quantification of the location uncertainty. The AE Sensors 1, 2, 3, and 4 had [x, y] locations of [145.0, 190.0], [145.0, 20.0], [20.0, 50.0], and [20.0, 220.0] mm, respectively, as seen in Fig. 2(c). As multiple sensing techniques were employed for damage monitoring in the CS, the AE sensor positions were selected through a trade-off with those sensor positions of the other techniques, in particular to maximize the monitoring region of both the AE and LWDS techniques. An amplitude threshold of 60 dB was set for capturing the hits to avoid the recording of noise signals. Only events localized within the AE sensor area are taken into account. The internal Vallen processor for planar positioning, which is based on Geiger's model, was used for localization (Broer

Table 1
The information of the composite specimens tested under C-C fatigue loading.

Year	Campaign 2019			Campaign 2020								
Name	L1–03	L1–05	L1–09	L1–49	L1–50	L1–51	L1–52	L1–54	L1–55	L1–56	L1–59	L1–60
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	At 0 cycles	At 0 cycles	At 0 cycles	After 5000 cycles	After 5000 cycles	After 5000 cycles	At 0 cycles	At 0 cycles	After 5000 cycles	After 5000 cycles	After 5000 cycles	After 5000 cycles
Size of disbond (mm)				15 × 20	20 × 20	20 × 25						
y-location of disbond (mm)				60	60	60						
Min Load (kN)	−6.5	−6.5	−6.5	−6.5	−6.5	−6.5	−6.5	−6.5	−6.5	−6.5	−6.5	−6.5
Max Load (kN)	−65	−65	−65	−65	−65	−65	−65	−65	−65	−65	−65	−65
Cycles (MTS)	152,458	144,969	133,281	48,702	65,500	94,431	368,558	510,961	226,356	756,226	110,137	170,884
Labeled Cycles	152,457	144,970	133,282	48,703	65,502	94,437	368,590	510,982	226,361	756,264	110,185	170,898
Error in cycles labeling	−1	1	1	1	2	6	32	21	5	38	48	14

Fig. 3. Workflow of the present work as a part of the general steps in prognostics.

et al., 2020). A filter was also used to exclude events with a position uncertainty greater than 50 mm. More detailed information for the applied localization method to AE data are described in subsection of Localizing Data. Six variables containing amplitude (A), rise time (R), energy (E), counts (CNTS), duration (D), and RMS have been extracted and recorded from AE events (see Fig. 4(a)).

4.2. Pre-processing

This section describes the pre-processing procedures, which include Labeling Cycles, Localizing Data, Windowing, and Missing Values. Another pre-processing step that should be conducted after feature extraction is a zero-mean normalization technique which will be addressed in Section 4.5.

• Labeling Cycles:

Labeling cycles on AE data is required for this study since a SSDNN with some hypothetical HIs as targets is proposed in order to construct HI, and the time (cycle) of each acquired AE event must be known in order to generate these targets. Due to the constraints of the MTS machine's output channels and the AE system's input channels and software, the AE system is unable to directly record cycles from the MTS machine. Nonetheless, since the AE system and MTS machine have been synchronized, and the AE system can import displacement and load values from MTS machine next to the other six variables from AE sensors, signal processing methods can approximate the cycle number of each hit, in which (cycle) that hit plus possibly more other hits occurred. Table 1 shows the number of cycles reported by the MTS machine (exact) and of the labeled cycles through the load signal (approximate), as well as the error between them. Given that the maximum error percentage is 0.044% (48/110137 for specimen 11), the estimated labeled cycles provided with the AE variables can be used.

• Localizing Data:

Geiger's method (Ge, 2003) was used to localize the AE data (Fig. 4(b)), and it allowed for planar localization of the AE events throughout the fatigue testing. This method assumes a constant wave velocity in all directions, which was determined using Hsu–Nielsen sources on pristine specimens (Broer et al., 2021). The wave speed was determined in both the x-direction (4423 m/s) and y-direction (6107 m/s), and the mean wave speed was then calculated as 5265 m/s. This was used as an input to Geiger's method to determine the planar location of the AE events. Since Geiger's method is a time-of-arrival approach, its application in anisotropic composite specimens can lead to errors in the AE event localization. The application of four AE sensors allows for the calculation of this position error, and a filter was implemented to exclude events with a position uncertainty larger than 50 mm. Lastly, events outside the AE sensor region are filtered.

• Windowing:

In the third step of pre-processing, as can be seen in Fig. 4(c), the signals (AE variables extracted from waveforms, including amplitude, rise time, duration, energy, counts, and RMS) are windowed for two reasons: one is that memorizing and analyzing all data from the beginning to the current time costs a tremendous computation time; another is that analyzing and comparing data at a single instant without taking into account nearby time steps is insufficient, especially for nonstationary signals. It should be noted that some SHM methods, like AE, do not record data points at a constant rate since they are passive methods and depend on the number of events occurring in the structure. For example, AE might measure 50 events in the first 10 s while it might measure 1000 events in the next 10 s. As can be seen in Fig. 5, the two main factors in the windowing process are the length of each window and the interval between two sequential windows, which are essential and important since they influence the final results and decisions. These factors might also be considered in a dynamic way rather than the static one and can be optimized as such. The windowing process for the current study was cycle-based, with a static length and interval of 500 cycles due to the natural interval of QS loads.

• Missing Values:

Since no events might have been recorded in a few intervals of 500 cycles due to the applied filters, there are missing values for those time windows. Because missing values have an impact on subsequent phases of the HI construction process, they should be eliminated or filled in, with the first option being taken for the windowed signals. Also, after the feature extraction step, some statistical features may be missed. For this step, linear interpolation is used to fill in the missing values.

4.3. Signal processing (SP)

A popular and logical way before FE is the transformation of data from the time domain into other beneficial domains like frequency and time–frequency domains. The foundation of SP in the frequency domain is frequency spectrum analysis which is based on Fourier Transform

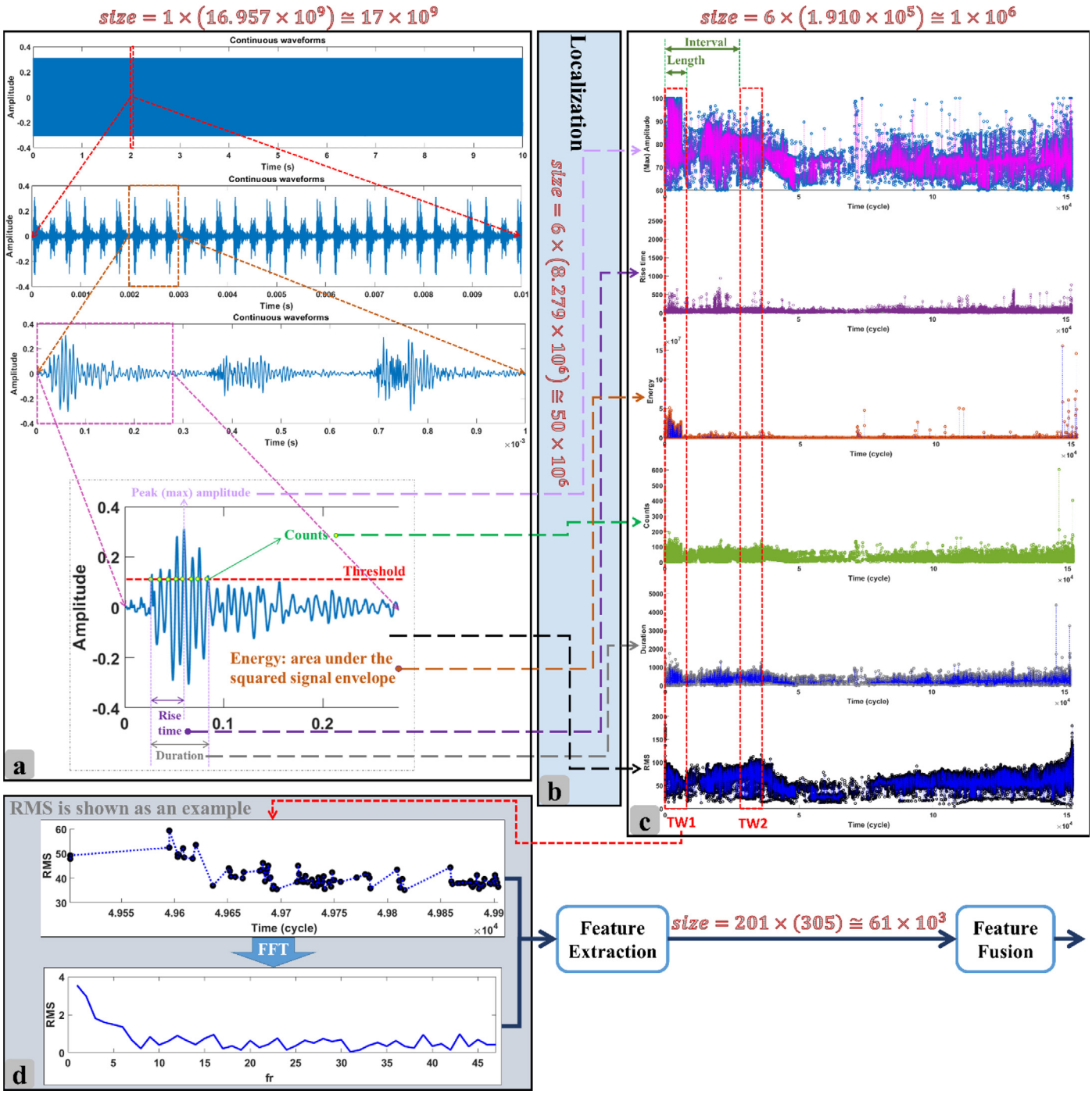

Fig. 4. Data reduction during (a) pre-processing (initial feature extraction), (b) localization, (c) windowing, and (d) signal processing.

(FT). To reduce computation time and improve the efficiency of the Discrete Fourier Transform (DFT) which is utilized for discrete signals rather than FT, the Fast Fourier Transform (FFT) is widely used to transform signals from the time domain into the frequency domain Cooley and Tukey (1965), and this SP method has been chosen in the present work.

It should be emphasized that the SP tasks are frequently attempted to be fulfilled entirely by a neural network. However, we argue that the mathematical gap between an artificial neuron and FT prevents a neuron from acting similarly to FT. A layer of artificial neurons, most likely with the same activation function, analyzes the inputs (x) multiplied by weights (Ws) with a single core activation function (like e^{-x}) resulting in e^{-Wx}, whereas FT has a multi-core activation function (e^{-Kx}) applied to inputs. In FT, the Ks are **explicitly** calculated and they are coefficients (frequencies) in the **perpendicular** spaces, whereas the Ws of ANN can be **implicitly** found by the back-propagation approach and there is no guarantee that the weights will become perpendicular to each other. As a result, it is worthwhile to analyze data using FT, which has an exact and fast solution and does not require any backpropagation process.

4.4. Feature extraction (FE)

In this step, features are extracted from multiple domains, including time and frequency domain. FE can also be carried out based on physical models which imply physical meaning but these model-based

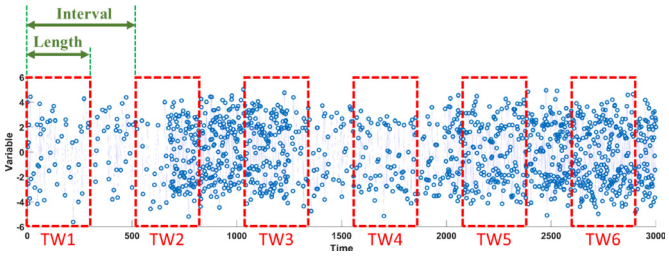

Fig. 5. Six time windows resulted by windowing process on artificial data with a constant length and interval.

features are limited. As a result, statistical parameters are extracted as features from various domains and employed in the following step (fusion).

The popular features in the time domain have been listed in Table A.1. However, as previously explained, the time domain is not sufficient to extract features as a HI or an element of HI, and additional statistical features should be extracted from the frequency domain. Furthermore, since an incomprehensible variation, especially in high frequency fluctuations, may not be detected in the time domain and instead it simply causes a spectrum line in the frequency domain, the frequency spectrum is more susceptible to incipient damages. This case is widely used in fault detection. The common features in the frequency domain have been listed in Table A.2.

As a result, 33 features are extracted from each of the 6 windowed signals (variables) of the AE data, including amplitude, rise-time, energy, counts, duration, and RMS. The broad features field has been expanded to include three additional possible useful features: cumulative Rise-time/Amplitude ratio, cumulative energy, and cumulative counts (Saeedifar and Zarouchas, 2020). The AE dataset yielded a total of 201 features ($6 \times 33 + 3$). It should be noted that FE procedure may also be regarded as a dimension reduction step (Fig. 4), as raw signals with thousands of data points within each window have been reduced to 201 data points. In fact, data with billions of records has been reduced to thousands.

4.5. Feature fusion (FF)

The extracted and selected features are fused in this step. The output of this step is called "Health Indicator (HI)" which will be imported in a prognostic model in order to predict RUL. It should be noted that not every feature can be considered as a prognostic parameter. There are several criteria that a feature or a fused feature should follow such that it can be accepted as a prognostic parameter. One of the main challenging issues of the current research is embedding these criteria in the fusion step and in the entire architecture of data processing and analysis as well as PHM. Thus, first, the aforementioned criteria are briefly introduced in Section 4.5.1, and then, a scenario to embed these criteria in the fusion model will be proposed in Section 4.5.2.

4.5.1. Criteria of prognostic parameters

In literature, a collection of metrics called Monotonicity (Mo), Prognosability (Pr), and Trendability (Tr) (Coble and Hines, 2009) have been suggested to examine whether an extracted and fused feature is acceptable as a HI or not.

a. Monotonicity (Mo)

The general increasing or decreasing pattern of a feature or generally a signal over time is expressed by Mo. According to several studies (Coble and Hines, 2009; Coble, 2010; Lei, 2016; Saidi et al., 2017), quantifying monotonic trend in a parameter can be identified in two manners, of which the details are provided next.

• Signum Formula or Sign Method:

The formula for this approach is:

$$Mo = \frac{1}{M}\sum_{j=1}^{M}\left|\sum_{k=1}^{N_j-1}\frac{sgn\left(x_j\left(k+1\right)-x_j\left(k\right)\right)}{N_j-1}\right| \tag{1}$$

where x_j is the vector of feature measurements on the j^{th} sample/system, M denotes the number of samples/systems monitored, and N_j denotes the number of measurements on the j^{th} sample/system. For this manner, two other similar equations (Eleftheroglou, 2020; Yue and Pilon, 2004) can also be used:

$$Mo = MK = \frac{1}{M}\sum_{j=1}^{M}\left|\sum_{i=1}^{N_j}\sum_{k=1,k>i}^{N_j}\left(t_k-t_i\right).sgn\left(x\left(t_k\right)-x\left(t_i\right)\right)\right| \tag{2}$$

$$Mo = MMK = \frac{1}{M}\sum_{j=1}^{M}\left|\frac{\sum_{i=1}^{N_j}\sum_{k=1,k>i}^{N_j}\left(t_k-t_i\right).sgn\left(x\left(t_k\right)-x\left(t_i\right)\right)}{\left(N_j-1\right)\sum_{i=1}^{N_j}\sum_{k=1,k>i}^{N_j}\left(t_k-t_i\right)}\right|.100\% \tag{3}$$

where $sgn(x)$ is:

$$sgn(x) = \begin{cases} -1 & if\ x<0 \\ 0 & if\ x=0 \\ 1 & if\ x>0 \end{cases} \tag{4}$$

t_k and t_i denote the times of measurements for $x\left(t_k\right)$ and $x\left(t_i\right)$, respectively. The difference between Mo in Eq. (1) with Mann–Kendall (MK) monotonicity in Eq. (2) is that the first one considers only the behavior of sequential data points and it does not measure the relation of data points with a time gap of more than one unit. Thus, Eq. (1) does not evaluate a more overall view and it is easily affected by noise. Nevertheless, since the output of the MK version is not normalized (e.g., a range of [−1,1]), the results do not have informative meaning (Eleftheroglou, 2020) and cannot be properly compared. In other words, the Mo weight is the same for all deterioration histories, regardless of time. Accordingly, the Modified Mann–Kendall (MMK) version in Eq. (3) is suggested (Eleftheroglou, 2020; Eleftheroglou et al., 2018) to measure the Mo of a function, which is used in the current study.

• Spearman's Rank Correlation Coefficient Method:

The Sign method of Mo is based on the number of sign change of the derivative, while the Spearman's Rank Correlation Coefficient Method is in accordance with correlation:

$$Mo = \frac{1}{M}\sum_{j=1}^{M}\left|\rho\left(rank\left(x_j\right),rank(t_j)\right)\right| \tag{5}$$

where t_j is the time-point vector that corresponds to the measurement vector x_j. $rank\,(x)$ computes the rank of matrix or vector x. This function employs a singular value decomposition (SVD) algorithm, which is more time-consuming than the previous approaches but still the most accurate. ρ is the correlation function, which can be, among others, Pearson's linear correlation coefficient (Benesty et al., 2009), Kendall Tau rank correlation coefficient (Kendall, 1948), and Spearman's rank correlation coefficient (Best and Roberts, 1975).

b. Trendability (Tr)

The term Tr refers to whether a parameter's decay histories (degradation) have the same underlying pattern for different samples or systems under monitoring. According to the literature (Coble and Hines, 2009; Coble, 2010; Lei, 2016; Saidi et al., 2017), the measure of similarity between trajectories of parameters can be expressed as:

$$Tr = \min_{j,k}\left|\rho\left(x_j,x_k\right)\right|,\quad j,\ k=1,\ 2,\ \ldots,M \tag{6}$$

The Pearson's correlation function is utilized in the current study, which is expressed as:

$$\rho\left(x_j, x_k\right) = \frac{cov\left(x_j, x_k\right)}{\sigma_{x_j}\sigma_{x_k}} \tag{7}$$

where cov is the covariance, σ_{x_j} and σ_{x_k} are the standard deviation of x_j and x_k, respectively. When the lengths of x_j and x_k vary, the shorter vector is resampled to equal the longer vector's dimension. To make this easier, their time vectors are first converted to percent lifespan, i.e., [0%, 100%].

Tr in Eq. (6) is based on correlation while Tr can also be expressed by (Khan and Yairi, 2018):

$$Tr = 1 - std(z_i), \quad z_i = \frac{no.\,of\,{}^{dy}/_{dt} > 0}{D-1} + \frac{no.\,of\,{}^{d^2y}/_{dt^2} > 0}{D-2} \tag{8}$$

where $y(t_i)$ denotes the value of measurement at the time of t_i, D is the number of measurements of the i^{th} degradation history. If each degradation pattern of the population for a parameter can be represented with the same functional form, that parameter attribute is trendable (Eleftheroglou, 2020). After testing and comparing different aforementioned approaches with artificial data, the former type (Eq. (6)) is selected to be uses in the present work because its outputs were more reasonable.

c. Prognosability (Pr)

The distribution of a parameter's failure (final) value is measured by Pr. The measure of the variability of parameters at failure (the final moment which is at the specimen's end-of-life) can be expressed as (Coble and Hines, 2009; Coble, 2010; Lei, 2016; Saidi et al., 2017):

$$Pr = exp\left(-\frac{\left(std_j x_j\left(N_j\right)\right)}{mean_j\left(\left|x_j(1) - x_j\left(N_j\right)\right|\right)}\right), \quad j = 1,\ 2,\ \ldots, M \tag{9}$$

d. Fitness

To simultaneously consider all the aforementioned criteria, an objective function called "Fitness" is used Eleftheroglou (2020):

$$Fitness = a \times Monotonicity_{HI} + b \times Prognosability_{HI} + c \times Trendability_{HI} \tag{10}$$

where a, b, and c are the control constants which determine how important each metric is compared to the others. These constants are often user-defined parameters (Eleftheroglou, 2020), which are usually assumed to be in the range of [0,1]. They are all considered one in this work, resulting in Fitness in a range of [0-3].

4.5.2. Semi-supervised criteria-based fusion neural network

In order to embed the Fitness function representing the criteria, different overall scenarios can be proposed. The simplest way is to rank the features according to their Fitness values. Although the feature with the highest rank can be considered as the HI, a threshold for the Fitness value can be set to accept more than one feature, and then, they can be considered as a set of HIs which are imported into a prognostic model. Otherwise, some simple methods such as a weighted averaging can be applied to the features filtered by the threshold, and finally, only one HI can be exported. Nevertheless, such approach might result in the overlooking of useful features since they do not fit the criteria. However, their combination may fulfill the intended specifications. Thus, as earlier mentioned, a fusion model based on DNN and Semi-Supervised Learning (SSL), rather than the predefined functions, is proposed in the current study.

SSL enables enormous volumes of unlabeled data to be exploited in conjunction with normally smaller labeled data sets (Van Engelen and Hoos, 2020). Unlabeled data can contribute to the formulation of a superior classifier or regressor, provided enough unlabeled data is available and certain assumptions about the distribution of the data are adopted. Semi-Supervised Classification (SSC) and Semi-Supervised Regression (SSR) are the key components of SSL (Kostopoulos et al., 2018), depending upon the type of the output variable.

In semi-supervised learning, there are several assumptions that define the forms of intended interaction (Chapelle et al., 2006). The most widely adopted assumptions are as follows Van Engelen and Hoos (2020):

(1) *Smoothness assumption*: two samples close to each other in the input space should result in close labels in the output space as well.
(2) *Low-density assumption*: the decision margin should not intersect across densely populated portions of the input space.
(3) *Manifold assumption*: the labels for sample points on the same low-dimensional manifold should be the same.

Most, if not all, semi-supervised learning algorithms are built on one or more of these assumptions. These assumptions are different definitions of the similarity between data points and their patterns (Van Engelen and Hoos, 2020).

The two most prevalent divisions in SSL are *transductive* and *inductive*, which are founded on the purpose of the training process. The former is merely concerned with providing labels for unlabeled data (not providing a model), whereas the latter constructs a classification or regression model that can be used to predict the label of unseen data points.

As adaptations of preexisting supervised algorithms, inductive learning algorithms that are called *intrinsically semi-supervised* (Van Engelen and Hoos, 2020) allow unlabeled data to be included in the objective function. They do not use any intermediary stages or supervised base learners, rather they directly optimize an objective function with components for labeled and unlabeled data. In general, these algorithms rely on one of the SSL assumptions, either explicitly or implicitly, and most semi-supervised neural networks rely on the smoothness assumption.

In the present work, a semi-supervised deep neural network (SS-DNN), by implicitly implementing the prognostic criteria as well as using the available EOL, has been proposed to construct HI by feature fusion. First, a hypothetical ideal HI function following the prognostic criteria is proposed and then used as a target for a supervised ANN to approximate the HI function (see Fig. 6). In this regard, two main questions arise:

1. How to select a suitable function to make an ideal HI conforming to the prognostic criteria?
2. What variable should be considered as the main variable of the ideal HI function?

which are discussed in the following subsection, *Selecting Hypothetical Ideal HI Function*. Then, *Multi-layers LSTM Network for Feature Fusion*, *Evaluation Metric*, *Hyperparameters Optimization and Model Validation* are addressed in the next three subsections.

a. Selecting Hypothetical Ideal HI Function

To create targets of the NN to predict HI, three aspects have been considered:

- The ideal HI function should best conform to all the criteria (Mo, Tr, and Pr).
- The simpler the ideal HI function, the simpler the NN model, and the faster its convergence.
- The ideal HI function should take into account the nonlinear behavior of damage propagation and accumulation in an engineering system, which is a composite structure in the current study.

The smoothness assumption of SSL is already taken into account when using an ideal HI function as a label generator that fulfills the prognostic criteria, i.e., if two bunches of extracted features at two different time steps are close to each other in the input space, their HIs (labels)

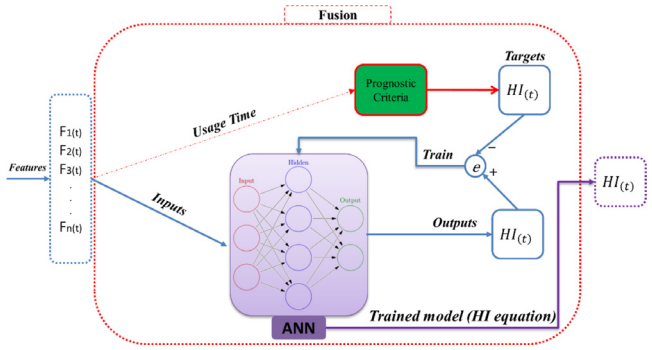

Fig. 6. Semi-Supervised Criteria-based NN Fusion to construct the HI from the features.

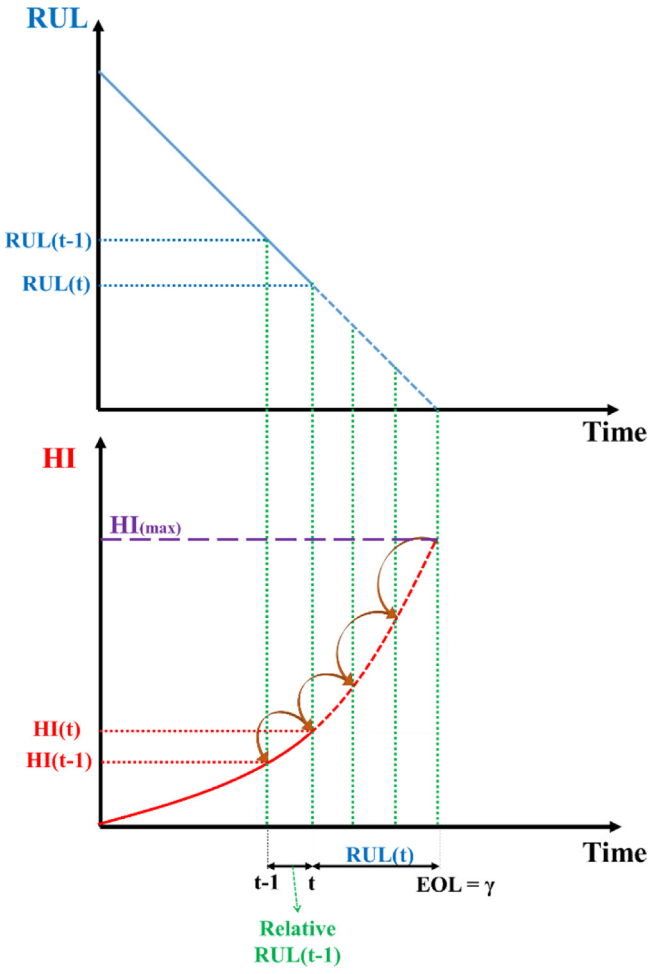

Fig. 7. Recursive reconstruction of the HI labels based on the relative RUL and the maximum HI at EOL to implement SSL.

are close to each other as well. Inversely, increasing the dissimilarity between two groups of extracted features at two separate time steps in the input space causes their HIs to move apart. In other words, the relative RUL between the former (t-1) and current (t) time windows, from which the features are extracted, is known. As a result, a direct relationship between the relative RUL and the relative HI variation (degradation) can be used to reconstruct the relative HI. By using the last relative RUL at the last time window before EOL (which is the only labeled data point concerning RUL for each sample) and considering a threshold as the maximum HI at EOL, all HI labels are recursively provided from the final failure at the EOL to the healthy state at the onset, yielding simultaneously prognosable behavior in HIs as the labels (see Fig. 7).

With the above-mentioned hypotheses in mind, four basic functions having high compatibility with the criteria are proposed and studied to select the best one: linear ($HI_t = \mathrm{t}$), quadratic polynomial ($HI_t = t^2$), natural logarithm ($HI_t = \ln(t)$), and exponential functions ($HI_t = \exp(t)$). These functions should be expressed in terms of usage time, which in this case is fatigue cycles. The functions should be normalized using the max–min normalization to adapt Pr as a recursive reconstruction process of HI. This normalization process is acceptable for hypothetical targets. In order to investigate the prognostic criteria, three different artificial specimens with variable lifetimes of 7, 4, and 10 time units (time step is 0.05) are considered, as shown in Fig. 8. All functions best match to Mo and Pr according to the calculated criteria shown in Fig. 9, however only linear and quadratic polynomial functions have the highest value (1) of Tr. As a result, the quadratic polynomial function is used to construct the targets since it takes into account damage propagation and accumulation nonlinearity. Therefore, the equation of ideal hypothetical HI for generating the targets is:

$$HI_{(t)} = normalize\left(t^2\right) = \frac{t^2 - HI_{min}}{HI_{max} - HI_{min}} = \frac{t^2 - t_0^2}{t_{EOL}^2 - t_0^2} \xrightarrow{t_0=0} HI_{(t)} = \frac{t^2}{t_{EOL}^2} \tag{11}$$

where t_0 and t_{EOL} are the usage times in terms of cycles at the beginning and the EOL, respectively. HI_{min} and HI_{max} are the minimum and maximum HI. The key point is that t_{EOL} is not available before the final failure.

Using the simulated labels, a loss function at the output layer can be defined, with the Mean-Squared-Error (MSE) being considered as the loss function in the current study. The half-mean-squared-error of the predicted outputs for each time step is the regression layer's loss function in sequence-to-sequence regression networks:

$$loss = \frac{1}{2S}\sum_{i=1}^{Q}\sum_{j=1}^{R}\left(T_{ij} - a_{ij}^{L}\right)^2 \tag{12}$$

where Q is the length of the sequence, R represents the number of responses, T_{ij} denotes target ($T_{ij} = \frac{t^2}{t_{EOL}^2}$), and a_{ij}^{L} represents the network's output (L refers to the last layer) for time step i and response j.

b. Multi-layers LSTM Network for Feature Fusion

Although the primary goal of this research is not to find and design the best and most generalized optimal ANN for fusing the extracted features, a suitable ANN architecture should be supplied and developed to demonstrate the feasibility of the suggested scenario. As a result, since the dataset for the examined composite structures is novel, and no previous study was performed on them, basic shallow architectures such as Multi-Layer Perceptron (MLP) were used to construct HI initially, before moving on to more complicated networks. In fact, each layer was added one at a time, the number of neurons at each layer was raised, and then the next layer was added. Various types of layers, such as the Fully Connected (FC) layer and Long Short Term Memory (LSTM), were tested in the meanwhile in order to achieve more satisfactory results. Fig. 10 depicts the finalized NN structure.

The NN architecture mainly consists of four types of layers including FC, Dropout, ReLU (Rectified Linear Units), LSTM, and Regression layers, of which the first four will be explained in the following subsections, respectively. Initially, it should be noted that the input features

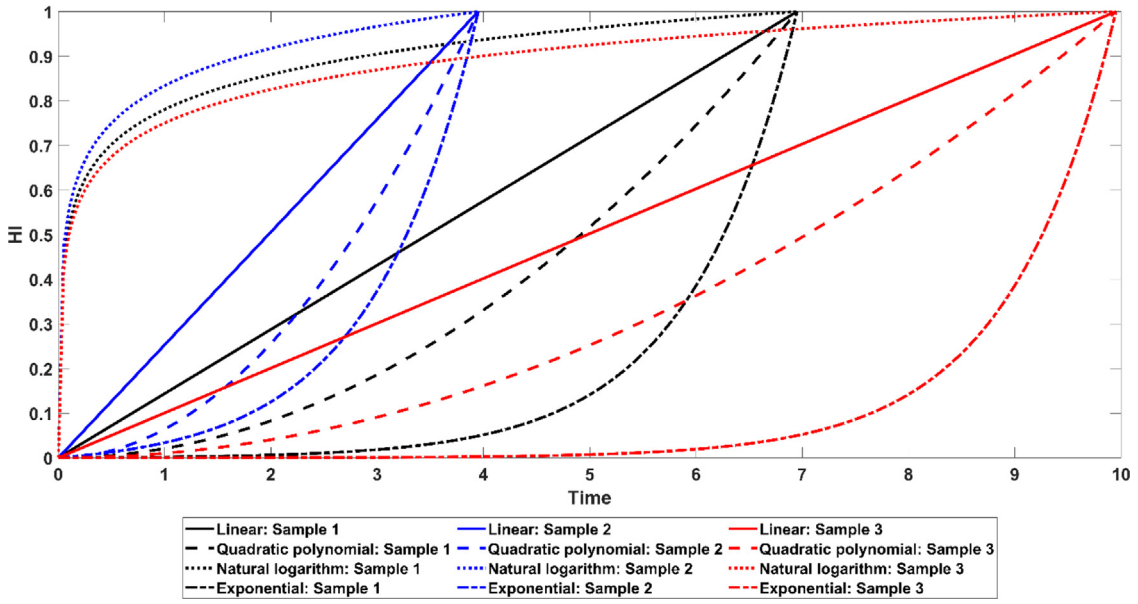

Fig. 8. Hypothetical HI functions for three artificial specimens with different lifetimes.

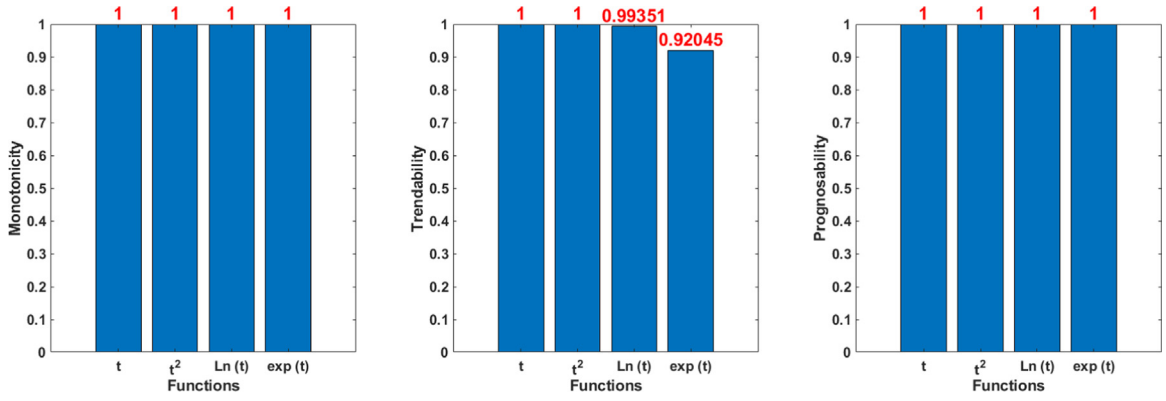

Fig. 9. Prognostic criteria of four hypothetical HI functions shown in Fig. 8.

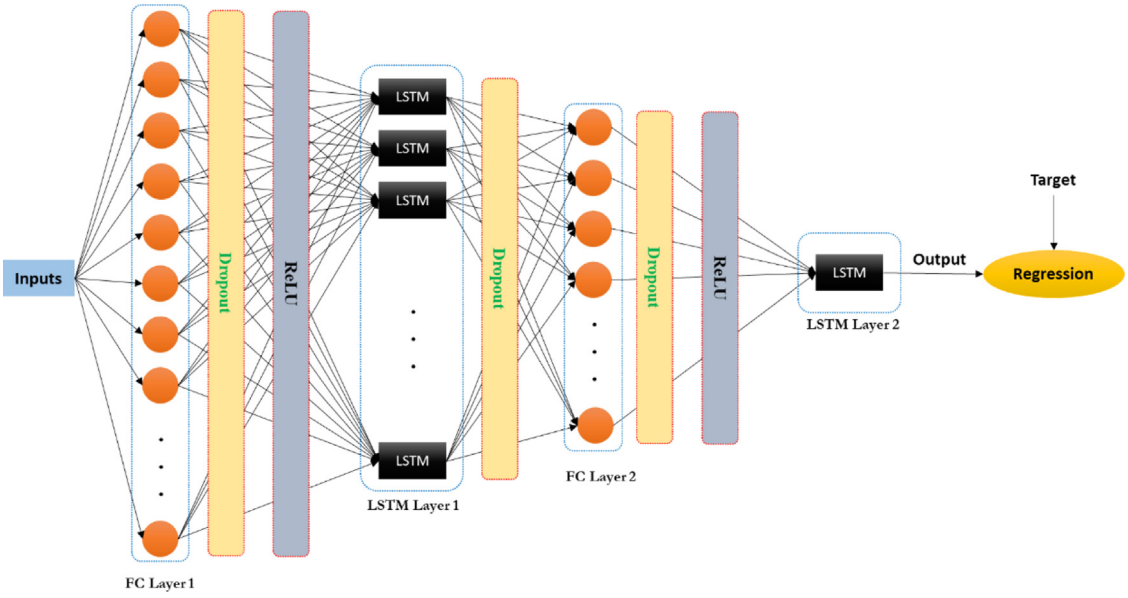

Fig. 10. Multi-layers LSTM network proposed for feature fusion.

were standardized before being imported into the network using a zero-mean normalization technique that used only the training dataset's mean value and standard deviation. The training data was also sorted by sequence length to reduce the amount of padding applied to the batches.

• Fully Connected (FC) Layer:

In FC layers, the neuron uses a weights matrix to apply a linear transformation to the input vector, which is called generalized linear layer ($z_j^l = \sum_k w_{jk}^l a_k^{l-1} + b_j^l$). A non-linear activation function σ is then used to apply a non-linear transformation to the product according to Eq. (13). If w_{jk}^l can be considered the weight for the link between the $(l-1)^{\text{th}}$ layer's k^{th} neuron to the l^{th} layer's j^{th} neuron, for a FC layer we have:

$$a_j^l = \sigma\left(z_j^l\right) = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) \tag{13}$$

where z_j^l and a_j^l denotes input and output of a desired FC layer l, respectively.

• Dropout Layer:

Dropout is a regularization strategy that prevents complicated co-adaptations on training data, thereby decreasing overfitting in artificial neural networks (Srivastava et al., 2014). The following equation is considered for the dropout layer:

$$\hat{w}_j = \begin{cases} 0, & with\ P(c) \\ w_j, & otherwise \end{cases} \tag{14}$$

where $\hat{w}_j$ is the diluted row and $P(c)$ is the probability c to remove a row in the weight matrix.

• Rectified Linear Units (ReLU) Layer:

ReLU is a type of activation function with a strong biological and mathematical foundation (Hahnloser et al., 2000). It consists of setting a threshold at 0:

$$ReLU\ (z) = z^+ = max(0, z) \tag{15}$$

• Long Short Term Memory (LSTM) Layer:

A memory cell (g), an input gate (i), an output gate (o), and a forget gate (f) compose an LSTM unit, which was developed to maintain a long-term record of sequential inputs by using the memory unit (Liang et al., 2021). As illustrated in Fig. 11, one input (x_t) and the previous hidden state (h_{t-1}) as well as the previous cell state (c_{t-1}) are used to formulate the hidden state (h_t) in the t$^{\text{th}}$ step in retaining information from the past as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{16}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{17}$$

$$g_t = tanh(W_g x_t + U_g h_{t-1} + b_g) \tag{18}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{19}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{20}$$

$$h_t = o_t \odot tanh(c_t) \tag{21}$$

where W_* and b_* stand for learnable weights and bias parameters, respectively. σ is sigmoid activation function and $\odot$ is the element-wise product. The hidden state h_t influences the production of the final output at any step t by accumulating information from previously processed features (Yang others, 2020), which could be referred to as damage accumulation and health degradation in the current study.

c. Evaluation Metric

Two performance metrics from the field of PHM are employed in this paper. The first evaluation criterion for the proposed scenario is the Root-Mean-Square Error (RMSE). The RMSE is expressed by Eq. (22):

$$RMSE = \sqrt{\sum_{i=1}^{Q} \frac{\left(T_i - a_i^L\right)^2}{Q}} \tag{22}$$

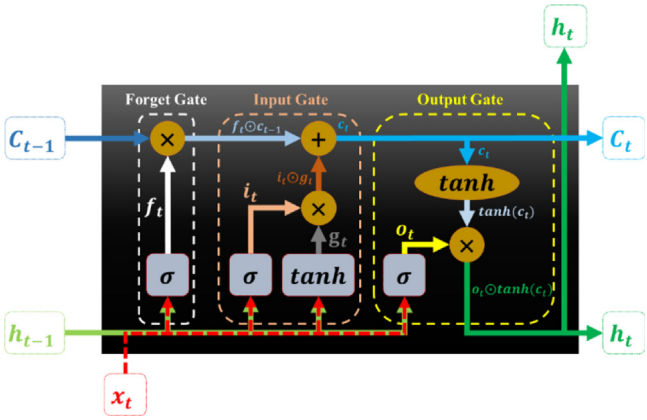

Fig. 11. LSTM cell's architecture.

where Q is the length of the sequence (HI), T_i and a_i^L denotes target and the network's output for time step i, respectively. This metric provides a single score for each CS's constructed HI. Furthermore, the $Fitness$ indicator, which encompasses three prognostic criteria outlined in Section 4.5.1, is the second evaluation metric.

d. Hyperparameters Optimization and Model Validation

After fixing an acceptable configuration of the neural network layers (Fig. 10), a Bayesian optimization algorithm (Snoek et al., 2012) was used to set the hyperparameters, including the number of neurons at each layer, Batch Size, and Dropout.

In this research, two main and trustworthy validation methods in the machine learning fields, which are Leave-One-Out Cross-Validation (LOOCV) and holdout validation methods (Arlot and Celisse, 2010), have been employed. First, using the holdout method with 10, 1, and 1 CSs as training, validation, and test datasets, respectively, the Bayesian optimization algorithm was used to find the top hyperparameters' sets and models (with the maximum RMSE over all CSs as the objective function which must be minimized). It should be noted that by considering the maximum RMSE of all CSs rather than other statistical parameters like the mean value of RMSE of them, the optimization algorithm attempts to simultaneously decrease the mean value and standard deviation of RMSE, which is more desirable. Then, the LOOCV method, having 11 and 1 CS as training and validation dataset, respectively, was applied to the top 10 models obtained by the holdout validation and Bayesian to check models' performance for the other folds, with 10 replications. Finally, the performance of these models will be described as a distribution with a mean expected error and a standard deviation.

5. Results and discussions

The deep learning framework and signal processing parts were developed using MATLAB R2021a; a high performance computing cluster (Beowulf style) with 12 processors on one node for the Bayesian optimization algorithm, and a laptop with an Intel Core i7-8665U CPU and 16 GB RAM for training the deep learning networks and the other parts (such as pre-processing and signal processing) were used. In this section, following the results of the holdout validation and the LOOCV, the best-proposed model will be discussed in comparison to the relevant literature in the subsection Discussion.

5.1. Holdout validation

First, the holdout method has been used to validate the model, with the first ten CSs for training, the 11$^{\text{th}}$ CS for validation, and the 12$^{\text{th}}$ CS for the test dataset.

An Adam optimizer (Kingma and Ba, 2014) was used to learn the deep learning model, with an initial learning rate of 0.005, a

Table 2
The top 10 hyperparameter sets (models) obtained by the Bayesian optimization algorithm and holdout validation with 11th CS as the validation and the 12th CS as the test dataset. RMSE is the maximum one over all CSs.

Model (rank)	Batch Size	Dropout	FCL1	LSTM1	FCL2	*RMSE*
1	4	0.3	110	154	50	*0.0829*
2	5	0.4	124	83	48	*0.0884*
3	5	0.5	201	79	36	*0.0983*
4	5	0.4	152	81	27	*0.1013*
5	5	0.5	41	142	43	*0.1026*
6	2	0	30	256	45	*0.1052*
7	5	0.4	124	56	48	*0.1055*
8	5	0.1	137	20	39	*0.1067*
9	5	0.4	161	92	48	*0.1084*
10	5	0.4	53	120	50	*0.1110*

learning rate drop factor of 0.2, a learning rate drop period of 5, and a gradient threshold of 1, which all have been selected after trial and error. Before each epoch, the training dataset was shuffled. Despite the fact that the maximum number of training epochs was set to 500, the network's output is based on the best validation loss, with the validation check frequency set to 30 iterations (number of trained batches) and validation check patience set to 6.

The Bayesian optimization algorithm was given 120 trials in parallel computing to optimize the hyperparameters. The number of neurons in the FC layer 1 and FC layer 2 as well as the number of units in the LSTM layer 1 have been allocated [1,201], [1,50], and [1,256], respectively, based on trial and error. It is worth noting that the LSTM layer 2 only contains one unit. For Dropout, the interval [0,0.5] quantized to 0.1 was also examined. Since the training dataset comprises ten samples, the interval [1,5] quantized to one has been explored for Batch Size. Since each set of Bayesian optimization final results is also dependent on the initial start points, the entire procedure was repeated several times. The top 10 hyperparameter sets (models) are presented in Table 2. As can be seen, the varied configurations have resulted in quite close RMSE ranged ~[0.08–0.11], which is the maximum RMSE over all CSs as the objective function of the Bayesian optimization algorithm.

Fig. 12(a) shows (merely as a case chosen to display the intuitive results) the constructed HIs by model 1, which is the first ranked, and their RMSE can be seen in Fig. 12(b). The RMSE for CSs 1, 11, and 12 slightly diverged from the mean value of RMSE for all CSs, which is 4%.

Since comparing the quality of the constructed HIs based solely on RMSE could not provide completely applicable information from a prognostic standpoint, the prognostic criteria Mo, Tr, and Pr, as well as their sum (Fitness), are shown in Fig. 13 for all the individual input extracted features as well as the HIs constructed by the model 1. The top four features with a Fitness score higher than 1.5 are features 185, 184, 88, and 183, respectively, which are (see Tables A.1 and A.2):

- feature 185: the 6th-order central moment of the RMS signal in the time domain
- feature 184: the 5th-order central moment of the RMS signal in the time domain
- feature 88: the variance of the Energy signal in the frequency domain
- feature 183: the 4th-order central moment of the RMS signal in the time domain

As can be seen, the high Fitness score of 2.891 for HIs, which is 77.3% higher than the best feature (1.630), demonstrates the high efficiency of the model 1 to construct HIs following the prognostic criteria. In fact, this Fitness improvement represents the performance of the proposed scenario and the whole developed algorithm because the model, the proposed DNN architecture, might still be enhanced by adding and/or changing characteristics such as the other types of layers, units, neurons, activation functions, and hyperparameters. However, as previously discussed, the main focus in the current research is how to implement the prognostic criteria in the process of HI construction. When the overall implementation methodology of the prognostic criteria has been validated, other enhancements like various optimization methods or DNN architectures can be studied in the main proposed roadmap. Nevertheless, the models are investigated in accordance with LOOCV in the next subsection, due to the shortcomings of the holdout validation in evaluating the generalization of the DNN models.

5.2. Leave-One-Out Cross Validation (LOOCV)

The ten models listed in Table 2 are tested with 10 repetitions on the 12 folds of LOOCV. It should be noted that the fold i refers to the fact that the i^{th} CS is the validation and the rest are training datasets. Fig. 14 shows the mean value and standard deviation of RMSE calculated over these repetitions for only the test dataset (e.g. for fold 1, CS 1 is the test dataset). In other words, the training (CSs) datasets were not taken into account during calculation of the mean value and standard deviation. For example, in the first fold of LOOCV, the 1st CS was considered the test, and CS 2–12 were considered for training, and when one model had completed training, the network was tested on the 1st CS. The mean value and standard deviation for that model and that fold (only for the test dataset which in this example is the 1st CS) were then calculated over ten repeats. This process has been performed for each of the 12 folds and each of the 10 best models from Table 2.

Fig. 14 demonstrates that some CSs, such as 3, 5, 7, 8, and 12, have better performance for all models; however, some, such as 1 and 6, suffer from randomness in the DNN algorithm, which is owing to the stochastic nature of ANN and randomness in the experimental data. Fig. 15 depicts a line plot of the mean value of RMSE with error (standard deviation) bars for all folds, illustrating a measurement of the generalization of the models. According to this figure, models 8, 2, and 7 are the best generalized ones with mean RMSE value of 0.121 ± 0.090, 0.133 ± 0.082, and 0.139 ± 0.085, respectively.

As previously stated, the Fitness scores could be more appropriate to report due to the deficiency in RMSE from a prognosis aspect. This can be performed in two ways:

(1) The Fitness score of the constructed HIs is measured for each replication; next, it is averaged across all replications; and finally, the findings for various models and folds are presented (Table 3 and Fig. 16).
(2) After all replications have been completed, the constructed HIs for each CS are averaged across all replications, and the Fitness score of the averaged HIs is calculated. Finally, the outcomes for various models and folds are provided (Table 4 and Fig. 18). This could be a case of models making **ensemble predictions**.

Each approach is discussed in more detail next. It should be noted that the Fitness scores for all extracted features remain unchanged and are identical to those in Fig. 13.

According to Table 3, the most challenging fold for the models to learn is the first one, with a mean Fitness value of 1.550 ± 0.216, and the other two worse folds are 10 and 2, with mean Fitness values of 1.780 ± 0.212 and 1.871 ± 0.139, respectively, and the rest have mean Fitness values of more than 2. The best generalized models are 2, 7, and 1 in order, with mean Fitness values of 2.360 ± 0.415, 2.342 ± 0.425, and 2.337 ± 0.427, respectively. To better compare different models and folds, the distribution of the average Fitness value of HIs for various folds and models can be seen in Fig. 16. Fig. 16(a) demonstrates that the models can appropriately construct HIs for folds 3, 5, 7, 8, 9, and 12 with a Fitness value greater than 2.575 ± 0.090 (fold 9). The remaining folds are affected by the model's low mean value or/and high variance. Model 6 has the lowest Fitness value averaged over all folds (2.230 ± 0.349), but it has the lowest variance (see Fig. 16(b)). The highest average Fitness value pertains to model 2 by which the

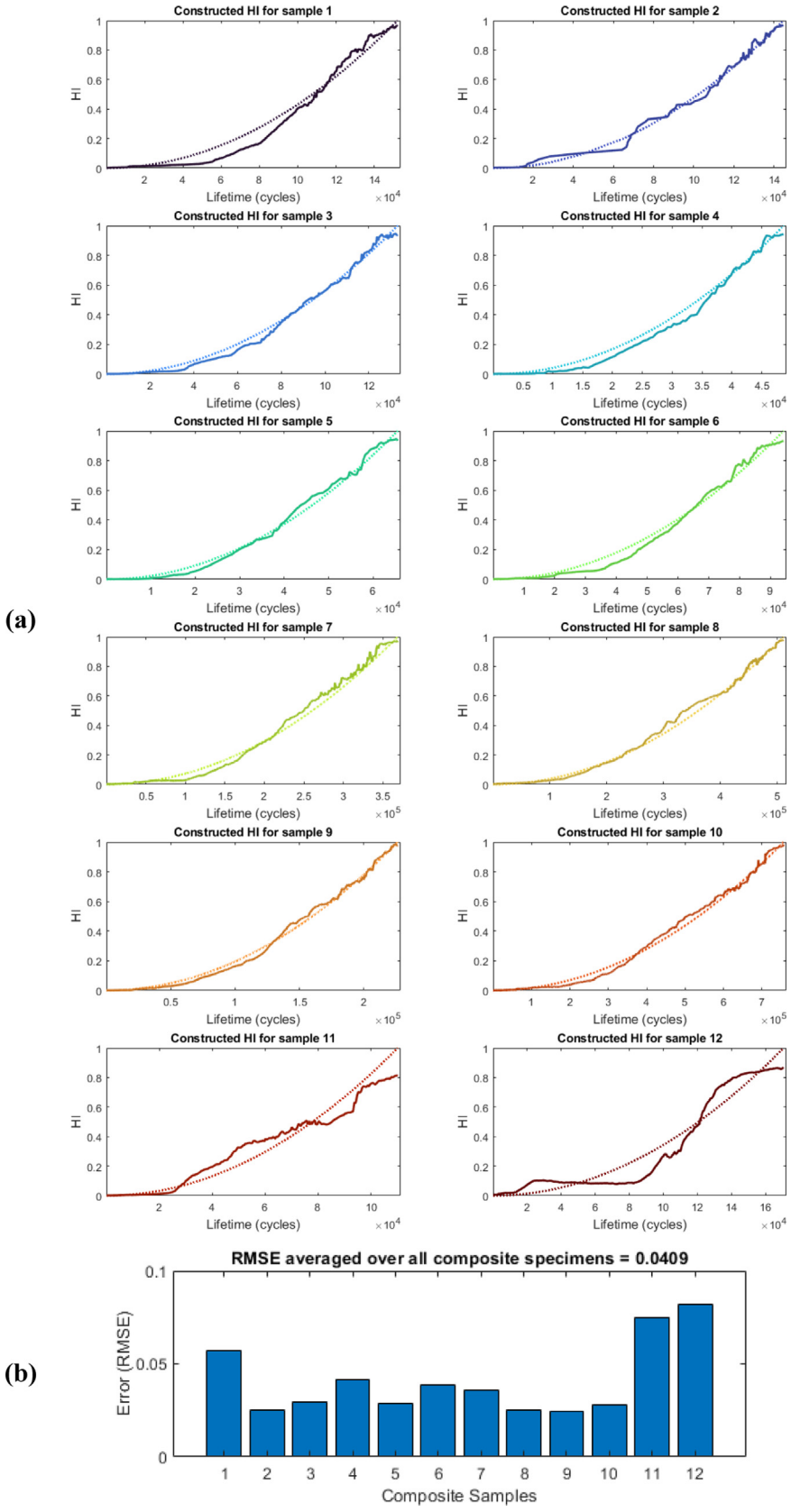

Fig. 12. (a) HIs constructed by model 1 and (b) their RMSE. The CS 11 and 12 are the validation and test datasets, and the rest are the training dataset. Dot lines are the target HIs.

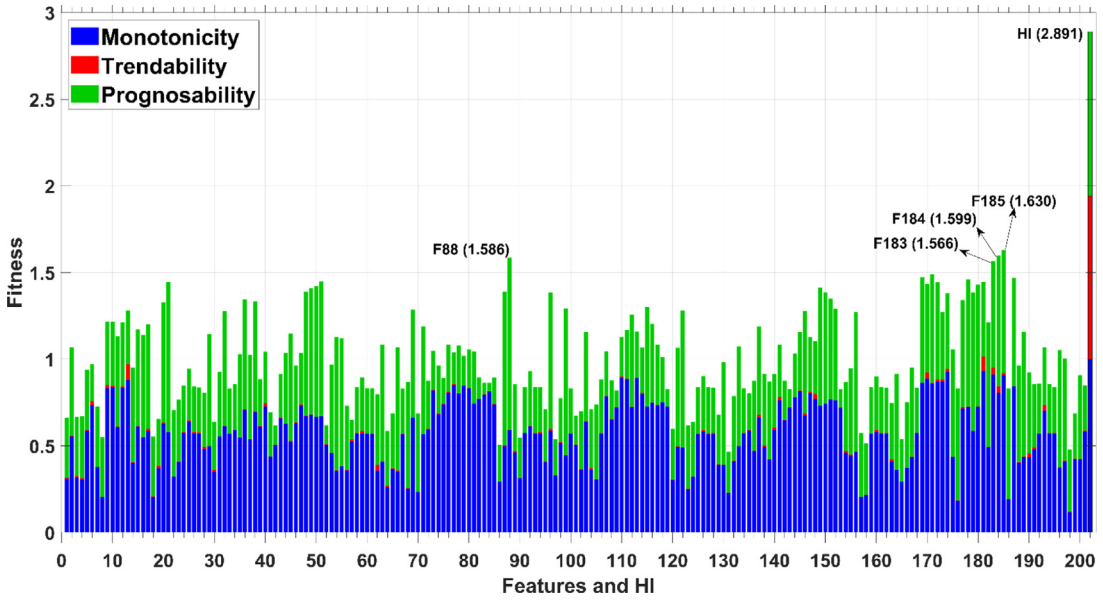

Fig. 13. The prognostic criteria for all 201 extracted features and the constructed HIs for model 1.

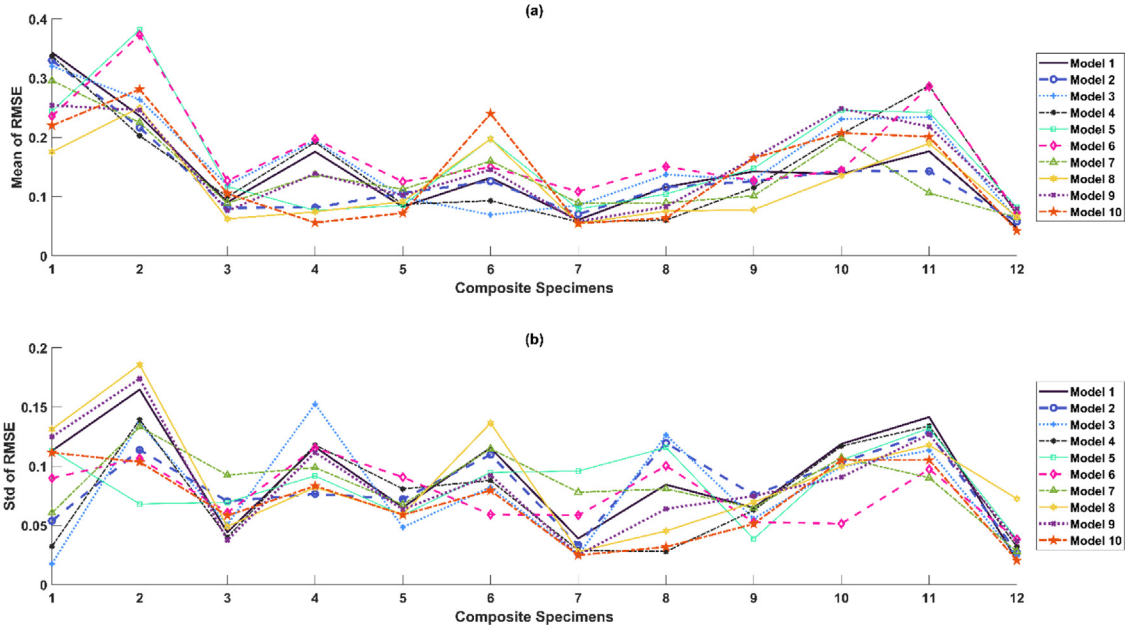

Fig. 14. (a) Mean value and (b) standard deviation of RMSE of test datasets (CS 1 to CS 12) over 10 repetitions for the top ten models.

constructed HIs in iteration 8 (best one) can be seen in Fig. 17. It is worth noting that the HIs for all CSs shown in this figure are from test datasets corresponding to relevant folds, and there are no constructed HIs from training datasets. Model 2 has not adequately learned CS 1, and also CSs 6, 10, 2, and 9 are not as qualified as the rest which have considerably good agreement.

So far, the prognostic criteria were averaged over HIs constructed by ten iterations. Hereinafter, the HIs constructed by ten iterations are averaged, and then, the average (ensemble) HIs are investigated to report the prognostic criteria (Fitness).

Table 4 shows that the first fold is again the most challenging for the models to learn, with a mean Fitness value of 2.019 ± 0.366, which is substantially better than what was reported in the previous state (Table 3), while the remainder have mean Fitness values of higher than 2.5. The best generalized models are 7, 8, and 5 in order, with mean Fitness values of 2.786 ± 0.144, 2.747 ± 0.146, and 2.729 ± 0.199, respectively. The average (ensemble) HIs obviously conform better to the prognostic criteria.

The distribution of the Fitness value of the average HIs can be seen in Fig. 18 to better compare various models and folds. Fig. 18(a) indicates that the models can construct HIs quite effectively for all folds except 1, 10, 6, and 2 when compared to the rest, in which the first fold with the lowest mean value and highest variance of Fitness is severe and distinguishable. The best Fitness value distribution pertains to model 7 by which the constructed average HIs (over all iterations) can be seen in Fig. 19(a). The discrepancy in deviation between the

Table 3
Averaged fitness scores over 10 repetitions of the constructing HI.

Model (rank)	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Fold 11	Fold 12
1	1.3715	1.9550	2.7984	2.0050	2.7291	2.2615	2.7084	2.5093	2.5791	2.0477	2.1968	2.8804
2	1.3663	2.0114	2.7561	2.4697	2.6075	2.1736	2.7587	2.5961	2.5629	1.9182	2.2626	2.8402
3	1.2923	1.8778	2.7210	1.9112	2.6438	2.5747	2.5844	2.6227	2.5624	1.5305	1.8487	2.7557
4	1.3048	1.9465	2.7622	1.8872	2.7150	2.3718	2.7805	2.7565	2.6879	1.7061	1.8616	2.7874
5	1.7903	1.5186	2.7209	2.5297	2.6217	1.7716	2.6271	2.6974	2.4861	1.6657	1.8323	2.7910
6	1.8221	1.7951	2.6852	1.9317	2.5055	1.9714	2.4565	2.4090	2.6288	2.0745	1.7897	2.6872
7	1.3730	1.9365	2.7382	2.3269	2.6166	2.1636	2.5459	2.5989	2.6861	1.7844	2.5063	2.8254
8	1.8142	1.9864	2.7427	2.4059	2.5146	1.7729	2.7520	2.6329	2.6575	2.0011	1.9362	2.6690
9	1.6320	1.9206	2.8524	2.1958	2.7016	2.2784	2.7852	2.7099	2.5107	1.4313	1.8569	2.7445
10	1.7323	1.7640	2.7064	2.4361	2.7218	1.6098	2.7767	2.7301	2.3921	1.6370	2.0289	2.8285
* "Green color → Red color" equalizes "Best result → Worst result"										***0***	***1.5***	***3***

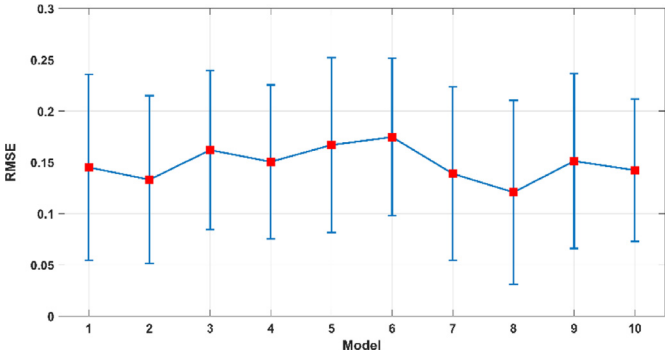

Fig. 15. Mean value of RMSE over all folds' tests (LOOCV) and after 10 replications, with error (Std) bars for the top ten models.

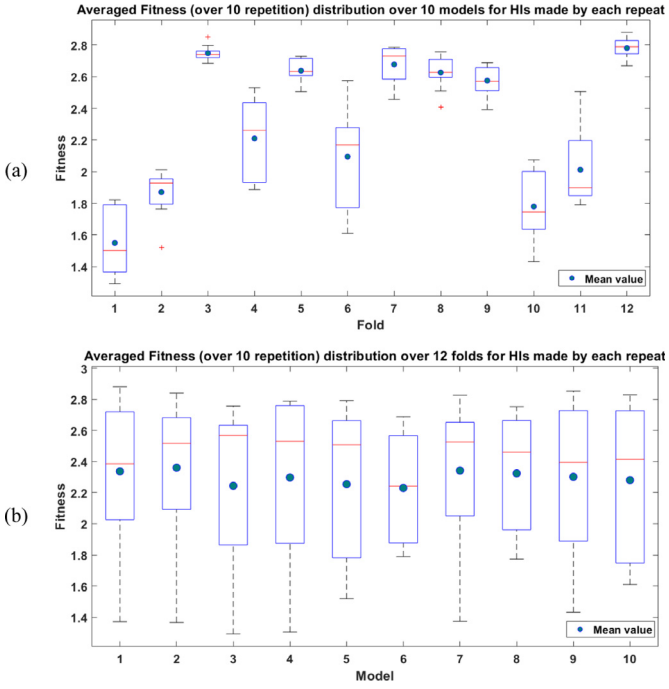

Fig. 16. The distribution of the averaged (across all replications) Fitness value of HIs for various (a) folds and (b) models.

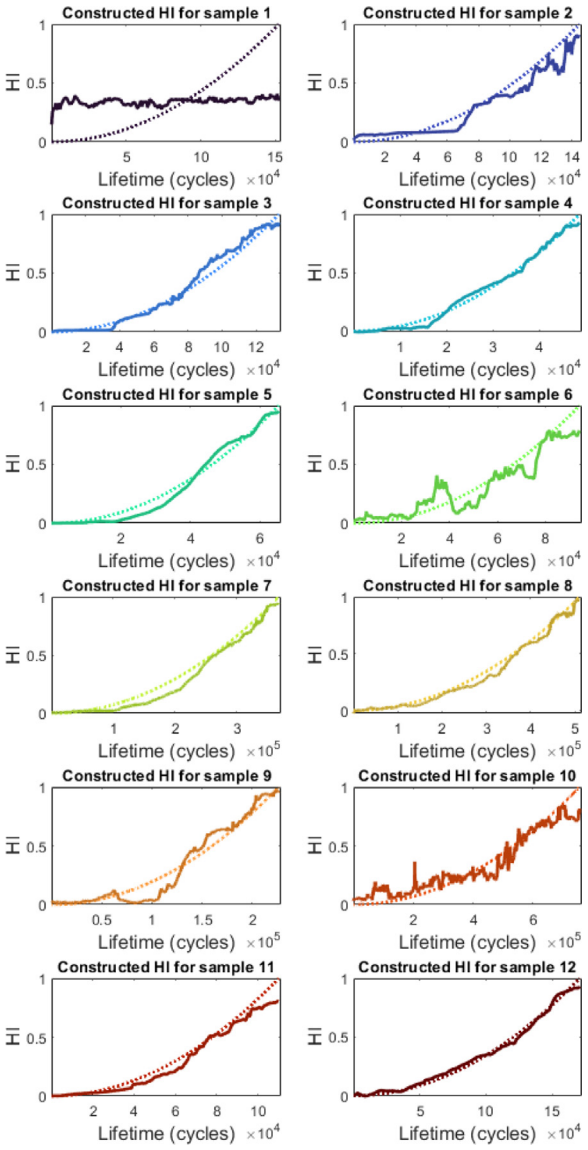

Fig. 17. HIs constructed by model 2 (in iteration 8 - best one). All shown HIs derive from test datasets matching to relevant folds, not training datasets, for all samples. Dot lines are the ideal HIs.

target, which is the ideal hypothetical HI, and the average constructed HI for the first fold (CS 1) is remarkable. Therefore, this fold containing its training dataset (CSs 2 to 12) has been shown in Fig. 19(b). Model 7 has obviously not learned the other CSs in the training dataset, let alone the test one, CS 1. It is possible that this is due to inappropriate training progress adjustments for this fold (e.g. validation check patience set to 6), demonstrating the limitations of the proposed DNN models, which can be improved in future work aimed at developing more generalized models and training progress for all folds. Nonetheless, Fig. 19(b) indicates that all CSs have comparable patterns to some extent, resulting in a fair Fitness value (2.453). With this in mind, while model 7 could not create HIs following the targets for the first fold (i.e., high RMSE), it could intelligently fuse the input features to

Table 4
Fitness scores of the averaged HIs over 10 repetitions.

Model (rank)	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Fold 11	Fold 12
1	2.0168	2.6074	2.9163	2.6504	2.8916	2.6407	2.8842	2.8194	2.8263	2.5998	2.8319	2.9233
2	1.4873	2.5428	2.9121	2.6661	2.8726	2.7153	2.8846	2.8594	2.8183	2.4931	2.8123	2.9051
3	1.5642	2.6752	2.9008	2.6609	2.8954	2.8413	2.8313	2.8513	2.8179	2.3088	2.7547	2.8882
4	1.7976	2.6279	2.9028	2.6517	2.8873	2.7437	2.8780	2.8416	2.8779	2.3911	2.7961	2.8702
5	2.6360	2.4077	2.8646	2.7460	2.8905	2.2435	2.9076	2.8856	2.7256	2.7750	2.8147	2.8545
6	2.3191	2.3126	2.8680	2.6015	2.8450	2.1449	2.7264	2.7789	2.8675	2.3156	2.6388	2.8118
7	2.4530	2.6766	2.9243	2.7039	2.8622	2.8455	2.8835	2.8528	2.8922	2.5723	2.8713	2.8924
8	2.4212	2.5125	2.8775	2.6465	2.8054	2.7307	2.8569	2.7969	2.8903	2.7093	2.8188	2.9009
9	2.2075	2.5598	2.9236	2.6181	2.8567	2.8423	2.8830	2.8545	2.7954	2.2688	2.8567	2.8252
10	2.1895	2.5205	2.8712	2.6471	2.8925	2.3005	2.8860	2.8649	2.6801	2.5750	2.8102	2.8892
* "Green color → Red color" equalizes "Best result → Worst result"										***0***	***1.5***	***3***

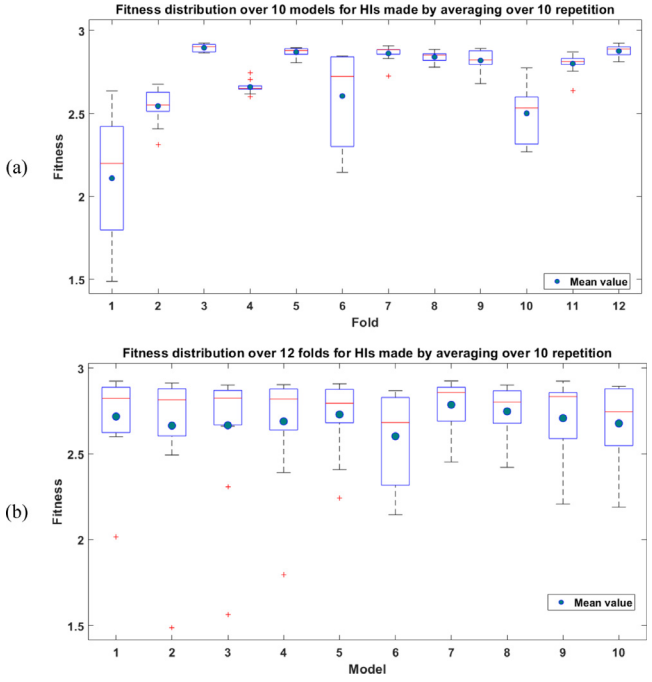

Fig. 18. The distribution of the Fitness value of averaged (across all replications) HIs for various (a) folds and (b) models.

produce an average HI that relatively matched the prognostic criteria (i.e., high Fitness), including Mo (almost increasing), Tr (almost same pattern), and Pr (almost from 0.1 to 0.4). In other words, the model could have discovered how to relate and fuse the features to create a HI with relatively the same pattern for all CSs.

5.3. Discussion

The best-generalized model according to the LOOCV is model 7, whose prognostic criteria in two statuses of single and ensemble models (ten parallel single models) are tabulated in Table 5. In addition, other literature relevant to composite structures has been presented, one of which (Ref. Yue et al. (2022)) worked with the ReMAP dataset, although it includes only 5 samples out of the 12 presented in the current paper. It should be noted that the more samples used, the lower the fitness score (quality) for HIs. Table 5 also shows the additional acoustic emission features that were considered as HI of the composite structure based on the literature, such as cumulated 1/A, energy, and rise-time/amplitude ratio (RA).

In terms of HI performance, the proposed ensemble model outperforms the other HIs proposed in the literature, according to Table 5. The prognostic criteria for the windowed AE features, particularly Tr, are extremely low. The maximum Mo is for the weighted HI derived from guided-wave (GW) data. Pr, on the other hand, is not as qualified as the proposed model in the current paper. It should also be highlighted that, even though both the AE and GW systems rely on acoustic and elastic waves within the structure, one is passive and the other is active. As a result, they measure different structural characteristics, resulting in a variety of data spaces. Therefore, the GW system may provide more informative data, whereas the physical model utilized in Ref. Yue et al. (2022) may not have been able to build its best HI. As a result, Table 5 compares the data's informativity as well as the proposed models for constructing HIs. Although only Mo has been quantitatively reported for the HIs obtained from DIC data, AE data, and the predefined function fusing DIC and AE data in Refs. Eleftheroglou et al. (2018) and Zarouchas and Eleftheroglou (2020), not only is Mo of the HIs developed in the current work slightly higher, but Tr and Pr are also superior based on qualitative comparison.

All in all, this work demonstrated that even when the true labels are unknown (as in HIs for composite structures) but the qualified pattern can be recognized (as in the prognostic criteria of Mo, Tr, and Pr), a model can still be built that aligns with the desired behavior using data-driven and artificial approaches.

6. Conclusions

A roadmap to construct an intelligent HI suitable for usage in prognostic models was proposed in the presented work. Following feature extraction with the purpose of dimensional reduction from the time and frequency domains of acoustic emission data captured during monitoring of single-stiffener composite panels, a feature fusion (FF) step based on semi-supervised learning was performed to ensure that the obtained HI after FF complies with the prognostic criteria. The quality of the constructed HI is measured and confirmed using these criteria, which include monotonicity, trendability, and prognosability. As a result, a semi-supervised deep neural network (SSDNN) was suggested to implicitly induce a multi-layer LSTM network that meets these criteria, as well as overcome the lack of labeled data and exploit unlabeled data to train models. Ten top models were selected from a Bayesian optimization algorithm applied to a holdout validation (ten CSs as training, one as validation, and one as a test), and they were then evaluated using the leave-one-out cross-validation (LOOCV) approach.

According to the holdout validation, the high Fitness score of 2.891 for HIs (maximum Fitness is 3) highlighted model 1's remarkable performance in constructing HIs based on prognostic criteria, which is 77.3% higher than the best feature (1.6303). Indeed, the proposed scenario's efficiency was reflected in this Fitness gain. Moreover, according to LOOCV, the best generalized models 7, 8, and 5 with the average (ensemble) HIs achieved mean Fitness values of 2.786 ± 0.144, 2.747 ± 0.146, and 2.729 ± 0.199, respectively. For the first fold, the difference in deviation between the target, which is the ideal hypothetical HI, and the average produced HI was considerable (e.g., model 7 has a mean RMSE value of 0.296 ± 0.061). It is likely due to unfit training progress adjustments for this fold (e.g. validation check patience set to 6), highlighting the proposed DNN models' limitations, which can be addressed in future work focused at designing more generalized models and training progress for all folds. Nonetheless, the results showed that

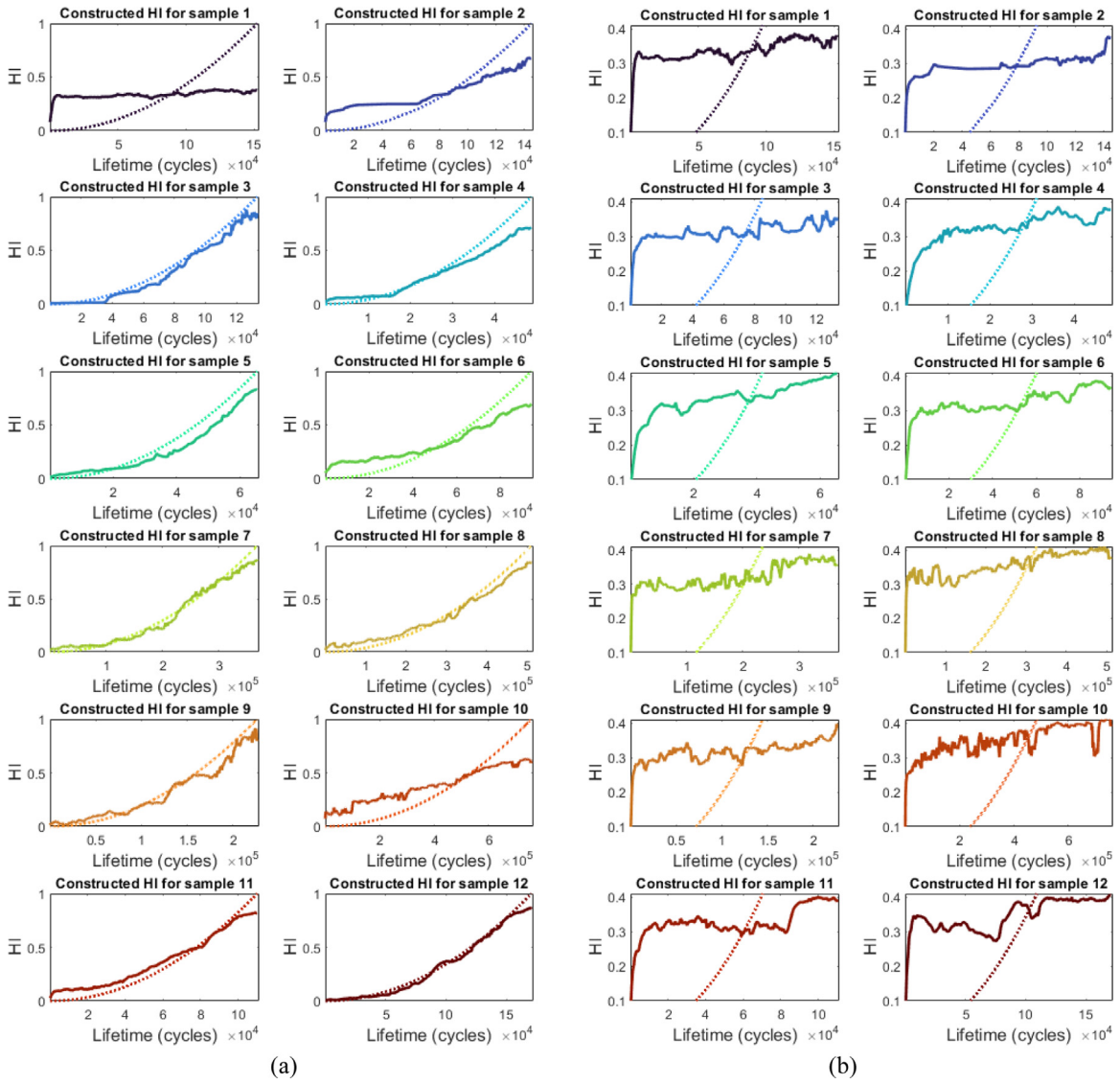

Fig. 19. Averaged HIs across all replications constructed by model 7 for (a) all folds (all samples are test datasets obtained from the relevant folds) and (b) fold 1 (only sample 1 is the test dataset (averaged), and the rest are the training ones (averaged)). Dot lines are the ideal HIs.

Table 5
Prognostic performance criteria of health indicators for composite structures.

	12 composite specimens of ReMAP					*5 composite specimens of ReMAP*					Eleftheroglou et al. [8, 24]*		
	Proposed models		According to [8, 24]	[38]	[39, 40]	Yue et al. [33]							
Prognostic Criteria	AE data					L-D** stiffness	GW data				DIC data	AE data	DIC & AE
	Single model (7)	Ensemble model (7)	1/A	E	RA		100 kHz	125 kHz	150 kHz	weighted	axial strain	1/A	Fusion
Monotonicity	0.87±0.15	0.99±0.01	0.57	0.5	0.59	0.66	1	1	0.92	1	0.94	0.91	0.98
Trendability	0.59±0.25	0.86±0.14	0	0	0	0.56	0.95	0.88	0.97	0.97	-	-	-
Prognosability	0.88±0.05	0.93±0.03	0.27	0.17	0.26	0.84	0.6	0.5	0.45	0.55	-	-	-
Fitness	2.34±0.45	**2.79±0.14**	0.84	0.67	0.85	2.06	2.55	2.38	2.34	2.52	-	-	-
"Green color → Red color" equalizes "Best result → Worst result"											***0***	***1.5***	***3***

** Different experiments to the current one.*
*** The stiffness values measured from load–displacement data.*

all CSs possessed patterns that were similar to some extent, resulting in a reasonable Fitness value (2.453 for model 7, indicating an 81.77% quality). While model 7 was unable to generate HIs that satisfied the first fold's targets (i.e., high RMSE), it was able to intelligently fuse the input features to yield an average HI that reflected the prognostic requirements (i.e., high Fitness). In other words, the model has figured out how to integrate and fuse the features to produce a HI with relatively the same pattern for all composite specimens, which can be utilized for more accurate and reliable prediction of EOL and decision-making. The prognostic criteria were implemented within the feature

Table A.1
Common statistical features in time domain.

No	Equation	Specific name	No	Equation	Specific name
1	$X_m = \frac{\sum_{n=1}^{N} x(n)}{N}$	Mean value	**9**	$X_{crest} = \frac{X_{peak}}{X_{rms}}$	Crest factor Daponte (2003)
2	$X_{sd} = \sqrt{\frac{\sum_{n=1}^{N} \left(x(n) - X_m\right)^2}{N-1}}$	Standard deviation	**10**	$X_{clearance} = \frac{X_{peak}}{X_{root}}$	Clearance factor
3	$X_{root} = \left(\frac{\sum_{n=1}^{N} \sqrt{\lvert x(n)\rvert}}{N}\right)^2$	Root amplitude	**11**	$X_{shape} = \frac{X_{rms}}{\frac{1}{N}\sum_{n=1}^{N} \lvert x(n)\rvert}$	Shape factor
4	$X_{rms} = \sqrt{\frac{\sum_{n=1}^{N} (x(n))^2}{N}}$	Root Mean Square (RMS) Paulter et al. (2004)	**12**	$X_{impulse} = \frac{X_{peak}}{\frac{1}{N}\sum_{n=1}^{N} \lvert x(n)\rvert}$	Impulse factor
5	$X_{rss} = \sqrt{\sum_{n=1}^{N} \lvert x(n)\rvert^2}$	Root-sum-of-squares (RSS) Paulter et al. (2004) *(The RSS level is also referred to as the 2-norm.)*	**13**	$X_{p2p} = \max(x(n)) - \min(x(n))$	Maximum to Minimum difference
6	$X_{peak} = max\lvert x(n)\rvert$	Peak	**14, 15, 16, 17**	$X_{k\_cm} = \frac{\sum_{n=1}^{N} \left(x(n) - X_m\right)^k}{N}$	Central moment for kth order *(In this context, the 3rd, 4th, 5th, and 6th are considered the features of 14, 15, 16, and 17, respectively.)*
7	$X_{skewness} = \frac{\sum_{n=1}^{N} \left(x(n) - X_m\right)^3}{(N-1)\,X_{sd}{}^3}$	Skewness	**18**	$X_{FM4} = \frac{X_{4\_cm}}{X_{sd}{}^4}$	FM4 *(close to Kurtosis)*
8	$X_{kurtosis} = \frac{\sum_{n=1}^{N} \left(x(n) - X_m\right)^4}{(N-1)\,X_{sd}{}^4}$	Kurtosis	**19**	$X_{med} = \frac{\sum_{n=1}^{N} t(n)}{N}$	Median

$x(n)$ denotes the signal sequence for $n = 1, 2, \ldots, N$.
N denotes the number of data points.
$t(n)$ denotes the moments of occurrence of $x(n)$.

Table A.2
Common statistical features in frequency domain.

No	Equation	Specific name	No	Equation	Specific name
20	$p_1 = X_{mf} = \frac{\sum_{k=1}^{K} s(k)}{K}$	Mean Frequency *(may indicate the vibration energy in the frequency domain, which represents the average of the amplitudes of all the frequencies.)*	**27**	$p_8 = \sqrt{\frac{\sum_{k=1}^{K} f_k{}^4 s(k)}{\sum_{k=1}^{K} f_k{}^2 s(k)}}$	
21	$p_2 = \frac{\sum_{k=1}^{K} (s(k) - p_1)^2}{K-1}$	*(same as variance)*	**28**	$p_9 = \frac{\sum_{k=1}^{K} f_k{}^2 s(k)}{\sqrt{\sum_{k=1}^{K} s(k) \sum_{k=1}^{K} f_k{}^4 s(k)}}$	
22	$p_3 = \frac{\sum_{k=1}^{K} (s(k) - p_1)^3}{K\left(\sqrt{p_2}\right)^3}$	*(same as Skewness)*	**29**	$p_{10} = \frac{p_6}{p_5}$	
23	$p_4 = \frac{\sum_{k=1}^{K} (s(k) - p_1)^4}{K {p_2}^2}$	*(similar to Kurtosis)*	**30**	$p_{11} = \frac{\sum_{k=1}^{K} (f_k - p_5)^3 s(k)}{K {p_6}^3}$	
24	$p_5 = X_{fc} = \frac{\sum_{k=1}^{K} f_k s(k)}{\sum_{k=1}^{K} s(k)}$		**31**	$p_{12} = \frac{\sum_{k=1}^{K} (f_k - p_5)^4 s(k)}{K {p_6}^4}$	
25	$p_6 = \sqrt{\frac{\sum_{k=1}^{K} (f_k - p_5)^2 s(k)}{K}}$		**32**	$p_{13} = \frac{\sum_{k=1}^{K} \sqrt{(f_k - p_5)} s(k)}{K \sqrt{p_6}}$	
26	$p_7 = X_{rmsf} = \sqrt{\frac{\sum_{k=1}^{K} f_k{}^2 s(k)}{\sum_{k=1}^{K} s(k)}}$		**33**	$p_{14} = \sqrt{\frac{\sum_{k=1}^{K} (f_k - p_5)^2 s(k)}{\sum_{k=1}^{K} s(k)}}$	

$s(k)$ denotes the spectrum for $n = 1, 2, \ldots, K$.
K denotes the number of spectrum lines.
f_k denotes the frequency value of the kth spectrum line.

fusion and HI construction framework for this advancement rather than being used solely as a quality measurement tool for HI.

CRediT authorship contribution statement

Morteza Moradi: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization. **Agnes Broer:** Investigation, Writing – review & editing. **Juan Chiachío:** Writing – review & editing, Supervision. **Rinze Benedictus:** Resources. **Theodoros H. Loutas:** Writing – review & editing. **Dimitrios Zarouchas:** Conceptualization, Writing – review & editing, Supervision.

Declaration of competing interest

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

Data availability

The data is publicly available.

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" and grant agreement No 769288 "ReMAP, Real-time Condition-based Maintenance for Adaptive Aircraft Maintenance Planning".

Appendix. Statistical features

The popular features in the time domain have been listed in Table A.1 by completing ones provided in Ref. Lei (2016).

The common features in the frequency domain have been listed in Table A.2 by completing ones provided in Ref. Lei (2016). In Table A.2, features $p2$–$p4$, $p6$, and $p11$–$p13$ characterize the convergence of spectrum power, which reflects frequency spectrum energy. Features $p5$ and $p7$–$p9$ can indicate a shift in the place of dominant frequencies in the frequency spectrum (Lei, 2016).

References

Al-Dulaimi, A., Zabihi, S., Asif, A., Mohammadi, A., 2019. A multimodal and hybrid deep neural network model for remaining useful life estimation. Comput. Ind. 108, 186–196.

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.

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Statist. Surv. 4, 40–79.

Benesty, J., Chen, J., Huang, Y., Cohen, I., 2009. Pearson correlation coefficient. In: Noise Reduction in Speech Processing. Springer, pp. 1–4.

Best, D., Roberts, D., 1975. Algorithm AS 89: the upper tail probabilities of Spearman's rho. J. R. Stat. Soc. Ser. C. Appl. Stat. 24 (3), 377–379.

Broer, A., Galanopoulos, G., Benedictus, R., Loutas, T., Zarouchas, D., 2021. Fusion-based damage diagnostics for stiffened composite panels. Struct. Health Monit. 14759217211007127.

Broer, A.A.R., Galanopoulos, G., Zarouchas, D., Loutas, T., Benedictus, R., 2020. Damage diagnostics of a composite single-stiffener panel under fatigue loading utilizing SHM data fusion. In: European Workshop on Structural Health Monitoring. Springer, pp. 616–625.

Cadini, F., Sbarufatti, C., Cancelliere, F., Giglio, M., 2019. State-of-life prognosis and diagnosis of lithium-ion batteries by data-driven particle filters. Appl. Energy 235, 661–672.

Chapelle, O., Schölkopf, B., Zien, A., 2006. Semi-Supervised Learning. Adaptive Computation and Machine Learning Series. The MIT Press, ed.

Coble, J.B., 2010. Merging data sources to predict remaining useful life–an automated method to identify prognostic parameters.

Coble, J., Hines, J.W., 2009. Identifying optimal prognostic parameters from data: a genetic algorithms approach. In: Annual Conference of the PHM Society, Vol. 1. no. 1.

Cooley, J.W., Tukey, J.W., 1965. An algorithm for the machine calculation of complex Fourier series. Math. Comp. 19 (90), 297–301.

Daponte, P., 2003. IEEE standard on transitions, pulses, and related waveforms.

Eleftheroglou, N., 2020. Adaptive prognostics for remaining useful life of composite structures.

Eleftheroglou, N., Loutas, T., 2016. Fatigue damage diagnostics and prognostics of composites utilizing structural health monitoring data and stochastic processes. Struct. Health Monit. 15 (4), 473–488.

Eleftheroglou, N., Zarouchas, D., Benedictus, R., 2020. An adaptive probabilistic data-driven methodology for prognosis of the fatigue life of composite structures. Compos. Struct. 245, 112386.

Eleftheroglou, N., Zarouchas, D., Loutas, T., Alderliesten, R.C., Benedictus, R., 2016. Online remaining fatigue life prognosis for composite materials based on strain data and stochastic modeling. In: Key Engineering Materials, Vol. 713. Trans Tech Publ, pp. 34–37.

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.

Ellefsen, A.L., Bjørlykhaug, E., Æsøy, V., Ushakov, S., Zhang, H., 2019. Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. Reliab. Eng. Syst. Saf. 183, 240–251.

Fink, O., Wang, Q., Svensen, M., Dersin, P., Lee, W.-J., Ducoffe, M., 2020. Potential, challenges and future directions for deep learning in prognostics and health management applications. Eng. Appl. Artif. Intell. 92, 103678.

Galanopoulos, G., Milanoski, D., Broer, A.A., 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.

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.

Ge, M., 2003. Analysis of source location algorithms: Part II. Iterative methods. J. Acoust. Emiss. 21 (1), 29–51.

Hahnloser, R.H., Sarpeshkar, R., Mahowald, M.A., Douglas, R.J., Seung, H.S., 2000. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. Nature 405 (6789), 947–951.

He, R., Tian, Z., Zuo, M.J., 2022. A semi-supervised GAN method for RUL prediction using failure and suspension histories. Mech. Syst. Signal Process. 168, 108657.

Hu, C., Youn, B.D., Wang, P., Yoon, J.T., 2012. Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. Reliab. Eng. Syst. Saf. 103, 120–135.

Huang, G., Hua, S., Zhou, Q., Li, H., Zhang, Y., 2020. Just another attention network for remaining useful life prediction of rolling element bearings. IEEE Access 8, 204144–204152.

Huang, C.-G., Yin, X., Huang, H.-Z., Li, Y.-F., 2019. An enhanced deep learning-based fusion prognostic method for RUL prediction. IEEE Trans. Reliab. 69 (3), 1097–1109.

Jiang, Y., Li, C., Yang, Z., Zhao, Y., Wang, X., 2021. Remaining useful life estimation combining two-step maximal information coefficient and temporal convolutional network with attention mechanism. IEEE Access 9, 16323–16336.

Kendall, M.G., 1948. Rank correlation methods.

Khan, S., Yairi, T., 2018. A review on the application of deep learning in system health management. Mech. Syst. Signal Process. 107, 241–265.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint.

Kostopoulos, G., Karlos, S., Kotsiantis, S., Ragos, O., 2018. Semi-supervised regression: A recent review. J. Intell. Fuzzy Syst. 35 (2), 1483–1500.

Lei, Y., 2016. Intelligent Fault Diagnosis and Remaining Useful Life Prediction of Rotating Machinery. Butterworth-Heinemann.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., Lin, J., 2018. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. Mech. Syst. Signal Process. 104, 799–834.

Li, T., Sbarufatti, C., Cadini, F., Chen, J., Yuan, S., 2022. Particle filter-based hybrid damage prognosis considering measurement bias. Struct. Control Health Monit. 29 (4), e2914.

Liang, F., Duan, L., Ma, W., Qiao, Y., Miao, J., Ye, Q., 2021. Context-aware network for RGB-D salient object detection. Pattern Recognit. 111, 107630.

Loutas, T., Eleftheroglou, N., Zarouchas, D., 2017. A data-driven probabilistic framework towards the in-situ prognostics of fatigue life of composites based on acoustic emission data. Compos. Struct. 161, 522–529.

Paulter, N.G., Larson, D.R., Blair, J.J., 2004. The IEEE standard on transitions, pulses, and related waveforms, Std-181-2003. IEEE Trans. Instrum. Meas. 53 (4), 1209–1217.

Pei, H., Si, X.-S., Hu, C., Li, T., He, C., Pang, Z., 2022. Bayesian deep-learning-based prognostic model for equipment without label data related to lifetime. IEEE Trans. Syst. Man Cybern.: Syst..

Peng, K., Jiao, R., Dong, J., Pi, Y., 2019. A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter. Neurocomputing 361, 19–28.

Qian, C., Westphal, T., Nijssen, R., 2013. Micro-mechanical fatigue modelling of unidirectional glass fibre reinforced polymer composites. Comput. Mater. Sci. 69, 62–72.

Qin, Y., Xiang, S., Chai, Y., Chen, H., 2019. Macroscopic–microscopic attention in LSTM networks based on fusion features for gear remaining life prediction. IEEE Trans. Ind. Electron. 67 (12), 10865–10875.

Saeedifar, M., Saleh, M.N., Nijhuis, P., de Freitas, S.T., Zarouchas, D., 2022. Damage assessment of a titanium skin adhesively bonded to carbon fiber–reinforced plastic omega stringers using acoustic emission. Struct. Health Monit. 21 (2), 407–423.

Saeedifar, M., Zarouchas, D., 2020. Damage characterization of laminated composites using acoustic emission: A review. Composites B 195, 108039.

Saidi, L., Ali, J.B., 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.

Seon, G., Makeev, A., Cline, J., Shonkwiler, B., 2015. Assessing 3D shear stress–strain properties of composites using digital image correlation and finite element analysis based optimization. Compos. Sci. Technol. 117, 371–378.

Snoek, J., Larochelle, H., Adams, R.P., 2012. Practical bayesian optimization of machine learning algorithms. Adv. Neural Inf. Process. Syst. 25.

Song, T., Liu, C., Wu, R., Jin, Y., Jiang, D., 2022. A hierarchical scheme for remaining useful life prediction with long short-term memory networks. Neurocomputing 487, 22–33.

Song, C., Liu, K., Zhang, X., 2017. Integration of data-level fusion model and kernel methods for degradation modeling and prognostic analysis. IEEE Trans. Reliab. 67 (2), 640–650.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15 (1), 1929–1958.

Van Engelen, J.E., Hoos, H.H., 2020. A survey on semi-supervised learning. Mach. Learn. 109 (2), 373–440.

Wang, C., Xiong, R., Tian, J., Lu, J., Zhang, C., 2022. Rapid ultracapacitor life prediction with a convolutional neural network. Appl. Energy 305, 117819.

Wen, P., Zhao, S., Chen, S., Li, Y., 2021. A generalized remaining useful life prediction method for complex systems based on composite health indicator. Reliab. Eng. Syst. Saf. 205, 107241.

Yang others, R., 2020. CNN-LSTM deep learning architecture for computer vision-based modal frequency detection. Mech. Syst. Signal Process. 144, 106885.

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.

Yue, S., Pilon, P., 2004. A comparison of the power of the t test, Mann–Kendall and bootstrap tests for trend detection/une comparaison de la puissance des tests t de Student, de Mann–Kendall et du bootstrap pour la détection de tendance. Hydrol. Sci. J. 49 (1), 21–37.

Zarouchas, D., Broer, A., Galanopoulos, G., Briand, W., Benedictus, R., Loutas, T., 2021. Compression Compression Fatigue Tests on Single Stiffener Aerospace Structures, V1 ed. DataverseNL.

Zarouchas, D., Eleftheroglou, N., 2020. In-situ fatigue damage analysis and prognostics of composite structures based on health monitoring data. In: Fatigue Life Prediction of Composites and Composite Structures. Elsevier, pp. 711–739.

Zhang, Y., Sun, J., Zhang, J., Shen, H., She, Y., Chang, Y., 2023. Health state assessment of bearing with feature enhancement and prediction error compensation strategy. Mech. Syst. Signal Process. 182, 109573.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.X., 2019. Deep learning and its applications to machine health monitoring. Mech. Syst. Signal Process. 115, 213–237.