JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 1, No. 1 (2023), pp. 17–40
DOI: 10.63646/jaiaa.2023.010102

Semi-Supervised AI Analytics for Prognostic Health Indicator Learning from Acoustic and Guided-Wave Sensor Streams

Wei Zheng[1]; Fang Liu[2]; Hao Chen[3, *]

[1] School of Mechanical Engineering, Shenyang Ligong University, Shenyang 110159, China
[2] Department of Control Science and Engineering, Hebei University of Technology, Tianjin 300401, China
[3] School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China
* Corresponding author: hchen@nuist.edu.cn

ARTICLE INFO
Received
October 18, 2022
Revised
December 21, 2022
Accepted
February 12, 2023
Available Online
March 30, 2023
DOI
10.63646/jaiaa.2023.010102

Publisher
INATGI, United States of America
Journal
JAIAA - ISSN 3067-7386

Abstract
Structural health monitoring of composite components used in aerospace, civil, and industrial machinery requires reliable prognostic health indicators (HIs) that track degradation continuously from initial healthy conditions to end-of-life. Constructing such HIs is exceptionally challenging because ground-truth damage labels are almost universally unavailable, damage mechanisms in heterogeneous materials are stochastic and multi-scale, and single-modality sensing inherently limits either temporal or spatial diagnostic resolution. This paper proposes a semi-supervised AI analytics framework that integrates passive acoustic emission (AE) sensor streams with active guided-wave (GW) interrogation data to derive high-quality fused HIs under severe label scarcity. The framework embeds three prognostic criteria—monotonicity (Mo), prognosability (Pr), and trendability (Tr)—directly into the learning process through physics-informed proxy targets and criteria-driven regularization, eliminating the need for manually annotated health trajectories. Dedicated intra-modality pipelines handle AE and GW separately before a late-fusion meta-learner combines unimodal HIs into a single prognostic index. Bayesian hyperparameter optimization and ensemble averaging stabilize the learned trajectories across random initializations. Strict leave-one-out cross-validation on a run-to-failure fatigue dataset of carbon-fiber-reinforced composite panels confirms that the proposed fused HI consistently achieves a cohort-level Fitness score above 2.70 (out of 3.0), representing improvements of 12–18 percentage points over single-modality baselines. Among the evaluated inter-modality fusion regressors, Gaussian process regression yields the highest test Fitness of 2.65 ± 0.10, demonstrating that complementary temporal (AE) and spatial (GW) information can be effectively unified within a semi-supervised prognostic learning paradigm.

Keywords: Acoustic emission; Guided waves; Health indicators; Semi-supervised learning; Structural health monitoring; Prognostics and health management; Sensor fusion; Composite structures

I. INTRODUCTION

Structural health monitoring (SHM) and prognostics and health management (PHM) are foundational disciplines for the safe and cost-effective operation of engineering assets across aerospace, civil infrastructure, and industrial machinery. At the heart of any PHM system lies the health indicator (HI): a scalar trajectory that compresses multi-sensor information into a compact representation of structural or mechanical condition, ideally tracking

degradation continuously from a pristine healthy state to end-of-life (EoL) (Farrar and Worden, 2007; Worden and Dulieu-Barton, 2004). Reliable HIs enable downstream tasks such as remaining useful life (RUL) estimation, maintenance scheduling, and anomaly detection. Constructing such indicators is, however, substantially harder than it may appear (Lei et al., 2018; Si et al., 2011).

The principal difficulty is the almost universal absence of ground-truth HI labels. In rotating machinery, standardized physical health indicators such as root mean square (RMS) vibration amplitude are well established because the dynamics amplify damage signatures in a predictable, largely monotonic manner. For composite structures subjected to cyclic fatigue loading, no universally accepted physical HI exists: progressive interacting damage mechanisms including fiber breakage, matrix cracking, delamination, and disbond initiation evolve stochastically across the material volume, and no single measurable quantity reliably captures this multi-scale process (Reifsnider and Gao, 1991; Degrieck and Van Paepegem, 2001; Talreja and Varna, 2016). This motivates the construction of virtual HIs (VHIs) that are engineered to satisfy prognostic objectives rather than to represent a directly measurable physical quantity.

Two complementary non-destructive sensing technologies are particularly well-suited to composite SHM. Acoustic emission (AE) is a passive method that records transient elastic waves generated by microscopic damage events such as matrix cracking and fiber debonding; it provides exceptional temporal sensitivity but limited spatial resolution and cannot directly quantify absolute damage severity at any given inspection time (Mba and Rao, 2006; Grosse and Ohtsu, 2008). Guided waves (GW), also known as Lamb waves in thin-plate structures, are an active technique in which piezoelectric transducers both excite and detect ultrasonic pulses; GW methods can estimate current damage dimensions and localize defects with good accuracy, but they acquire only discrete snapshots at each inspection cycle and carry weaker information about damage growth history between snapshots (Su and Ye, 2009; Mitra and Gopalakrishnan, 2016; Raghavan and Cesnik, 2007). These complementary characteristics create a compelling motivation for multi-modal sensor fusion: AE contributes high-resolution temporal information about damage initiation and propagation dynamics, whereas GW provides high-fidelity spatial snapshots of the structural state at each inspection epoch (Khaleghi et al., 2013; Lahat et al., 2015).

Despite the intuitive appeal of multi-modal fusion, combining AE and GW into a unified prognostic framework introduces several algorithmic challenges. The two modalities operate on entirely different timescales: AE records events continuously at rates of thousands of hits per second, whereas GW snapshots are acquired at intervals of hundreds to thousands of fatigue cycles. Synchronizing these asynchronous streams into a coherent timeline suitable for joint learning is non-trivial (Castanedo, 2013). Furthermore, the high dimensionality of AE feature spaces—commonly exceeding 200 engineered statistical and spectral descriptors—demands effective dimensionality reduction without sacrificing prognostically relevant variation. On the machine learning side, the absence of labeled training data rules out standard supervised approaches and demands a learning strategy that can nonetheless produce HIs conforming to known physical constraints on health degradation (Van Engelen and Hoos, 2020; Zhu and Goldberg, 2009; Chapelle et al., 2006).

Semi-supervised learning (SSL) offers a principled route to this challenge. By replacing unobservable ground-truth labels with physics-informed proxy targets derived from known EoL behavior, and by embedding prognostic quality criteria directly into the training objective, it is possible to guide the model toward monotonic, prognosable, and trendable HI trajectories without any manual annotation. This idea has recently gained traction in PHM for rotating machinery (Zhao et al., 2019; Lei et al., 2020) but has not been systematically applied to the problem of multi-modal composite structure monitoring under the strict constraints of leave-one-out generalization testing.

This paper makes the following contributions. First, we propose a complete end-to-end semi-supervised AI

analytics pipeline that integrates AE and GW sensor streams through hierarchical intra- and inter-modality fusion. Second, we embed monotonicity, prognosability, and trendability criteria directly into the hyperparameter optimization objective via Bayesian optimization, enabling unsupervised, semi-supervised, and hybrid training strategies within the same framework. Third, we systematically benchmark eight inter-modality fusion regressors—including Gaussian process regression, gradient boosting, support vector machines, multilayer perceptrons, and LSTM—under strict leave-one-out cross-validation on a run-to-failure fatigue dataset of carbon-fiber-reinforced polymer (CFRP) composite panels. Fourth, we introduce test-only variants of the prognostic criteria that prevent inflated cohort-level scores from masking poor generalization to held-out units. The remainder of this paper is organized as follows: Section II reviews theoretical foundations; Section III describes the proposed framework; Section IV details the experimental setup; Section V presents and discusses results; and Section VI concludes.

II. THEORETICAL BACKGROUND

A. Health Indicators and Prognostic Criteria

A health indicator is a scalar time series $x_j(t_i)$ that tracks the degradation state of unit j at observation time t_i, spanning from a healthy baseline to EoL. Three widely used criteria assess HI quality: monotonicity (Mo), prognosability (Pr), and trendability (Tr) (Saxena et al., 2008; Coble and Hines, 2011). Monotonicity quantifies whether the HI trajectory consistently increases or decreases across the service life—higher values near EoL are expected for a damage indicator. Prognosability captures how concentrated EoL HI values are across multiple similar units, normalized by the average magnitude of HI change over life; higher Pr implies that all units reach a consistent degraded state at failure, which is essential for reliable RUL prediction. Trendability measures whether the shapes of degradation trajectories are consistent across units, computed as the minimum pairwise correlation among all unit trajectories; higher Tr indicates that the degradation mode is reproducible across manufacturing and loading variability. A combined Fitness score aggregates all three: Fitness = Mo + Pr + Tr (range [0, 3]), with higher values indicating better prognostic utility (Liao and Kottig, 2014).

To mitigate the well-known problem of overly optimistic training-cohort Fitness values, test-only variants (Fitness_Test) are computed by excluding the held-out test unit from Pr and Mo calculations, ensuring that reported performance reflects true out-of-unit generalization. This distinction is particularly important in small-cohort settings, such as the five-unit GW dataset used in this study, where training-cohort means are highly sensitive to the inclusion of any single unit (Si et al., 2011; An et al., 2013).

B. Acoustic Emission for Composite Monitoring

Acoustic emission is generated when localized energy is rapidly released within a material under stress, producing transient elastic waves that propagate to the surface and are detected by broadband piezoelectric sensors. In composite laminates, AE sources include matrix cracking, fiber pullout, fiber fracture, delamination initiation, and disbond propagation. Each source type produces characteristic waveforms distinguishable by features such as peak amplitude, rise time, duration, energy, counts, and frequency content, yielding high-dimensional feature vectors per detected hit (Grosse and Ohtsu, 2008; Sause et al., 2012). Because AE records events in real time at the rate they are generated, cumulative AE statistics carry rich information about damage accumulation history, making them naturally suited to time-series learning architectures that exploit sequential ordering.

A persistent challenge in AE-based prognostics is the stochasticity of composite damage: identical specimens tested under nominally identical conditions can produce very different AE activity patterns, reflecting variability in fiber alignment, resin distribution, and void content. Additionally, external noise sources—mechanical vibration, electromagnetic interference, fixture friction—can contaminate AE signals with spurious hits. Effective AE frameworks

must therefore include hit localization to discard geometrically implausible events, appropriate windowing to aggregate hit statistics over meaningful time intervals, and robust dimensionality reduction to extract prognostically relevant information from the high-dimensional feature space (Mba and Rao, 2006; Huang et al., 1998; Bengio et al., 2013).

C. Guided Waves for Composite Monitoring

Guided wave methods excite and receive ultrasonic Lamb waves through bonded piezoelectric transducers. In the pitch-catch configuration, one actuator transmits a tone-burst at a selected frequency while multiple receivers capture the transmitted wave over different paths. The presence of damage scatters, attenuates, or mode-converts the guided wave, producing changes in the received signal relative to a healthy baseline. Damage index features extracted from the signal envelope, frequency content, and path-specific characteristics provide quantitative state estimates at each inspection epoch (Giurgiutiu, 2007; Mitra and Gopalakrishnan, 2016; Lim et al., 2014). The dispersive and multimodal nature of Lamb waves in composite laminates complicates interpretation, since multiple modes coexist at most frequencies and their characteristics depend on stacking sequence, ply properties, and local stiffness.

A practical advantage of GW for prognostics is history-independence at each inspection: the received signals can be compared to a pristine baseline without requiring continuous monitoring between snapshots. This property also represents a limitation—GW cannot capture the real-time dynamics of damage initiation that AE records continuously. In multi-modal integration, GW snapshots serve as anchor points that constrain the absolute health state while AE trajectories fill in the temporal evolution between anchors (Raghavan and Cesnik, 2007; Su and Ye, 2009).

D. Semi-Supervised Learning for PHM

Semi-supervised learning addresses settings in which only a small fraction of training examples have labels, by leveraging the structure of the unlabeled data to improve model performance. In the PHM context, a fully inductive variant of SSL is more appropriate than transductive SSL: the trained model must generalize to entirely new units not seen during training, rather than making predictions for specific unlabeled examples in the training set (Van Engelen and Hoos, 2020; Chapelle et al., 2006). Several inductive SSL approaches have proven effective in PHM, including pseudo-labeling—where confident model predictions on unlabeled data are treated as provisional labels—and consistency regularization—where the model is encouraged to produce similar outputs for perturbed versions of the same input (Lee, 2013; Tarvainen and Valpola, 2017; Berthelot et al., 2019; Sohn et al., 2020).

The physics-informed proxy label approach adopted in this work is a specialized form of SSL in which the proxy targets are not model predictions but physics-derived surrogates for the expected HI shape. Specifically, a quadratic function of normalized time-to-failure is used as the proxy target, consistent with the empirical observation that composite damage accumulates nonlinearly and accelerates toward EoL (Coble and Hines, 2011; Liao and Kottig, 2014). The model is trained to minimize the discrepancy between its output and this proxy target while simultaneously optimizing the prognostic quality criteria through Bayesian hyperparameter tuning. This dual-objective strategy encourages the model to produce HIs that are both proxy-consistent and prognostically meaningful, even when the proxy scale and the true HI scale do not match (Sikorska et al., 2011; Goebel and Bonissone, 2005).

III. PROPOSED FRAMEWORK

A. Overall Architecture

The proposed semi-supervised AI analytics framework consists of three main stages: (1) dedicated intra-modality pipelines for AE and GW; (2) a synchronization and resampling procedure that aligns the two modality timelines into a common representation; and (3) an inter-modality late-fusion meta-learner that estimates a single fused HI from the paired unimodal HI estimates. Figure 1 presents the complete architecture. All components operate under

the inductive semi-supervised learning paradigm: proxy labels derived from end-of-life physics replace unavailable ground-truth HIs, and prognostic criteria are embedded into both the training loss and the hyperparameter optimization objective (Chapelle et al., 2006; Van Engelen and Hoos, 2020).

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 1, No. 1 (2023), pp. 17–40
DOI: 10.63646/jaiaa.2023.010102

Figure 1.

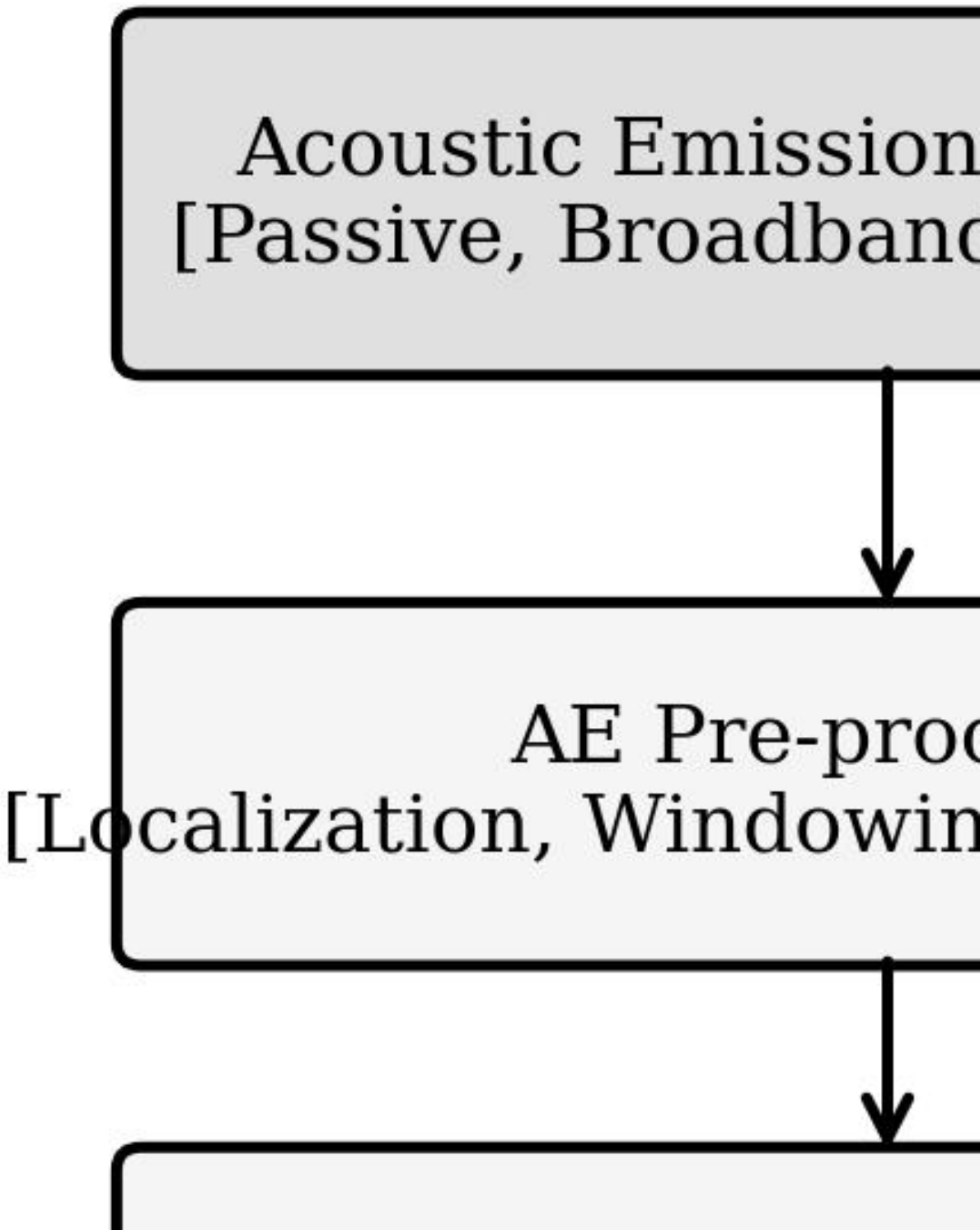


Figure 1. Multi-level intra- and inter-modality sensor fusion framework for semi-supervised health indicator learning. Dedicated data-driven pipelines handle AE (left column) and GW (right column), and a late-fusion meta-learner combines the unimodal HIs into a single prognostic index.

A key design choice is the separation between unimodal pretraining and inter-modality fusion. Unimodal AE and GW models are each trained and cross-validated independently on their respective full cohorts, which are larger than the overlap cohort of specimens instrumented with both sensing systems. The pretrained unimodal models are then frozen and used to generate HI estimates for the smaller fusion cohort, which serves as the training set for the meta-learner. This transfer learning strategy (Pan and Yang, 2010) preserves the representational power of each modality-specific pipeline while managing the data scarcity in the fusion cohort.

B. Intra-Modality AE Frameworks

Three complementary AE frameworks are implemented to capture diverse aspects of the temporal AE data. All three share a common preprocessing chain: six standard per-hit features (amplitude, rise time, energy, counts, duration, and RMS) are extracted from the continuous AE record; hit localization using the Geiger arrival-time method discards events with position uncertainty exceeding 50 mm; and a sliding time window converts the irregular event stream into fixed-resolution feature matrices suitable for sequential learning. The frameworks diverge in their signal processing and modeling choices.

AE-Framework 1 (FFT-SSLSTM) applies fast Fourier transform spectral features to augment the time-domain descriptors, yielding 201-dimensional feature vectors per window. A compact semi-supervised LSTM network maps these sequences directly to a proxy-target HI trajectory. The loss function minimizes the mean squared error between the model output and the quadratic proxy target. Hyperparameters are tuned via Bayesian optimization with a validation objective combining root mean squared error and cohort-level Fitness (Snoek et al., 2012; Shahriari et al., 2016). Ten independent training runs from random initializations are averaged using simple average ensembling (SAE) to reduce seed sensitivity (Dietterich, 2000).

AE-Framework 2 (PCA-2S-ML-PBO) separates temporal and spatial information into a two-stage pipeline. After adaptive per-unit online standardization, principal component analysis (PCA) reduces the 201-dimensional feature vector to ten principal components (Jolliffe, 2002). A time-independent model (TIM)—a four-layer fully connected network—first maps the compressed features to an intermediate HI estimate. A lightweight LSTM time-dependent model (TDM) then refines this estimate by modeling sequential dependencies. Hyperparameter optimization uses a physics-based Bayesian objective (PBO) that explicitly penalizes low Fitness in addition to proxy target error, ensuring that the optimization actively favors prognostically meaningful configurations (Snoek et al., 2012). Weighted average ensembling (WAE) with weights determined by cohort Fitness improves stability further (Polikar, 2006; Freund and Schapire, 1997).

AE-Framework 3 (CEEMDAN-SSEDL) replaces FFT spectral analysis with complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), which decomposes each AE window into intrinsic mode functions (IMFs) that adapt to local signal characteristics (Bengio et al., 2013; Hinton and Salakhutdinov, 2006). This yields 504-dimensional feature vectors per window, capturing both time-scale separation and transient non-stationarities characteristic of progressive composite damage. A semi-supervised encoder-decoder deep learning (SSEDL) architecture maps these features to HI trajectories under the same proxy-label and criteria-regularization scheme as the other frameworks. A deep learning ensemble (DLE) combining ten diverse initializations with learned blending weights serves as the final AE-F3 predictor (Hochreiter and Schmidhuber, 1997; LeCun et al., 2015).

C. Intra-Modality GW Framework

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 1, No. 1 (2023), pp. 17–40**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2023.010102**

The GW framework operates on discrete inspection snapshots acquired at 5,000-cycle intervals. At each inspection, one of eight surface-bonded PZT sensors acts as an actuator and transmits tone-bursts at six excitation frequencies (50, 100, 125, 150, 200, and 250 kHz), while the remaining seven sensors serve as receivers, covering 56 distinct actuator-receiver paths per inspection. After baseline subtraction using the healthy-state reference signal, damage index features—including cross-correlation coefficients, signal energy ratios, peak amplitude changes, and time-of-flight shifts—are extracted from each path-frequency combination, yielding high-dimensional GW feature vectors per snapshot.

A semi-supervised deep neural network (GW-SSL-DNN) maps these snapshot features directly to an HI estimate, training on proxy labels derived from the quadratic EoL profile and optimized via an unsupervised Fitness-based objective that does not require proxy alignment for the absolute HI scale. Homogeneous intra-modality fusion across excitation frequencies within the GW framework is achieved by ensembling models trained on each frequency subset individually, with WAE weights determined by the per-frequency Fitness score. This frequency-level ensembling captures complementary damage sensitivity at different scales—low frequencies (50–125 kHz) being more sensitive to large-area delaminations, and high frequencies (150–250 kHz) being more sensitive to local matrix cracking and fiber breakage (Giurgiutiu, 2007; Lim et al., 2014).

Figure 2 illustrates the conceptual contrast between conventional supervised learning, which requires unavailable ground-truth labels, and the proposed inductive SSL paradigm, which substitutes physics-informed proxy targets and criteria-driven regularization to guide the learning process without annotation.

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 1, No. 1 (2023), pp. 17–40
DOI: 10.63646/jaiaa.2023.010102

Figure 2. Supervised

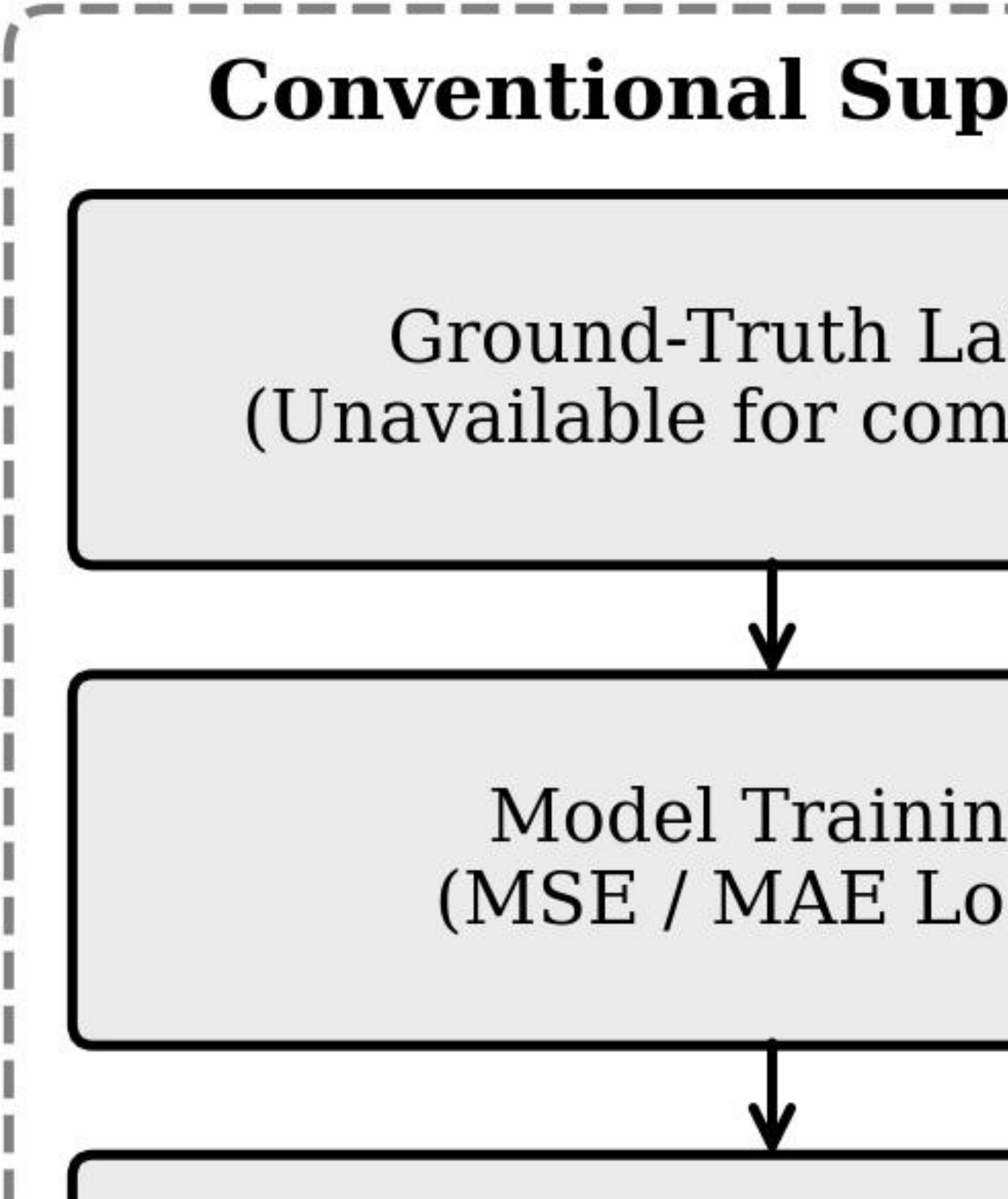


Figure 2. Conceptual comparison of conventional supervised learning (left) and the proposed inductive semi-supervised learning paradigm (right). In the absence of ground-truth health labels, physics-informed proxy targets combined with prognostic criteria regularization enable label-free HI construction.

D. Bayesian Hyperparameter Optimization

All frameworks use Bayesian optimization (BO) to tune hyperparameters under objectives that directly encode prognostic criteria. For AE-F1, BO minimizes the maximum RMSE between model-generated HIs and proxy targets across all training units. For AE-F2, a composite physics-based objective (PBO) combines normalized proxy error and negative Fitness, ensuring that the optimizer cannot increase proxy fidelity at the cost of prognostic quality. For the GW framework, BO minimizes negative Fitness directly on the validation set, making the optimization purely unsupervised with respect to proxy targets. This spectrum of optimization strategies—semi-supervised (PBO), unsupervised (Fitness-only), and hybrid—provides a comprehensive assessment of how different supervision levels affect HI quality (Snoek et al., 2012; Shahriari et al., 2016; Brochu et al., 2010).

$$L\_PBO = L\_reg + \beta \cdot L\_Crit \;\; where \;\; L\_Crit = (3 - (Mo + Pr + Tr)) / 3$$

In the above, L_reg is the normalized proxy RMSE computed on the validation set, L_Crit is the normalized criteria loss computed over training and validation units jointly, and the hyperparameter β controls the relative weighting. The default setting $\beta = 1$ is used in all AE-F2 experiments, but sensitivity analysis confirms stable behavior across the range $\beta \in [0.5, 2.0]$.

E. Timeline Synchronization and Inter-Modality Fusion

AE provides a continuous record at the event timescale (sub-second), while GW provides snapshots at 5,000-cycle intervals. To enable inter-modality fusion, the AE-derived HI trajectory is resampled at each GW inspection epoch by interpolating the AE HI at the corresponding normalized life position. This alignment yields synchronized pairs (HI_AE(t_i), HI_GW(t_i)) at each inspection time t_i, which together constitute the input to the meta-learner.

Eight meta-learners are evaluated: Gaussian process regression (GPR) with a Matérn kernel, least-squares boosting (LSBoost), decision tree (Tree), stepwise linear regression (LR_S), robust linear regression with bisquare weights (LR_R), support vector machine with a radial basis function kernel (SVM), multilayer perceptron (MLP), and LSTM. GPR is particularly attractive because it provides principled uncertainty quantification in addition to point estimates, which is valuable for maintenance decision-making under epistemic uncertainty (Rasmussen and Williams, 2006; Mao et al., 2014). LSBoost and MLP provide strong nonlinear regression capacity, while the linear models serve as interpretable baselines. All meta-learners are trained on pairs from the non-test units within the fusion cohort and evaluated on the held-out test unit.

IV. EXPERIMENTAL SETUP

A. Dataset Description

Experiments are conducted on run-to-failure compression-compression (C-C) fatigue data from the H2020 ReMAP project. The specimens are single-stiffener CFRP composite panels with an IM7/8552 carbon-fiber-reinforced epoxy skin and a T-shaped stiffener, following an Embraer aerospace design. The skin layup is [±45/0/45/90/-45/0]_S and the stiffener layup is [±45/0/±45]_S. Panels were loaded at 2 Hz with an R-ratio of 10 (compressive cycling). To introduce realistic variability, controlled 10 J impact events were applied at different locations on selected specimens, and some panels contained pre-existing disbond defects of different sizes and locations introduced via PTFE inserts during manufacturing. In total, 14 units were tested to failure; valid AE records were available for 12 units and valid GW records for 5 units. Only 3 units were instrumented with both sensing systems, making the inter-modality fusion task particularly challenging from a data scarcity perspective.

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed
RESEARCH ARTICLE
Vol. 1, No. 1 (2023), pp. 17–40
DOI: 10.63646/jaiaa.2023.010102

Table I. Key Specifications of Selected CFRP Panel Specimens from the ReMAP Fatigue Campaign

Parameter	L1-03	L1-05	L1-09	L1-23	L1-49	L1-55
Load Range (kN)	−65 / −6.5	−65 / −6.5	−65 / −6.5	−65 / −6.5	−60 / −6	−65 / −6.5
Lifetime (cycles)	152,458	280,098	144,969	133,281	438,000	170,884
AE Monitoring	✓	✓	✓	✓	✓	✓
GW Monitoring	✓	–	✓	–	✓	–
Disbond Size (mm)	None	None	None	30×30	15×20	None
Impact Energy (J)	10	10	10	0 (none)	10	10
Impact Cycle	0	0	0	–	5,000	5,000

Note: Check marks (✓) indicate sensor system deployed; – indicates not instrumented; L1-03, L1-05, L1-09 from 2019 campaign; L1-23, L1-49, L1-55 from 2020 campaign.

B. Sensor Configuration and Signal Acquisition

For AE monitoring, four Vallen Systeme GmbH VS900-M broadband sensors covering 100-900 kHz were bonded to the panel surface using a standard coupling agent and clamped holders. The Vallen AMSY-6 acquisition system recorded all channels simultaneously at a sampling rate of 2 MHz with a threshold of 35 dB(AE). Hit-based feature extraction applied the Geiger two-dimensional planar localization algorithm, and events with localization error exceeding 50 mm were discarded to ensure that only damage-related events within the specimen gauge region were retained.

For GW monitoring, eight PZT discs (Noliac NCE51, 10 mm diameter, 0.5 mm thick) were bonded in two rows spanning the stiffener region. The National Instruments PXI system drove each actuator sequentially with windowed tone-burst excitations at six center frequencies (50, 100, 125, 150, 200, 250 kHz), producing 2-cycle packets with a 20-cycle wait for wave field decay. The full set of 336 actuator-receiver-frequency combinations was recorded at each inspection epoch, covering both symmetric and asymmetric propagation modes. Baseline signals acquired before fatigue loading were subtracted from all subsequent records to isolate damage-induced changes (Giurgiutiu, 2007; Mitra and Gopalakrishnan, 2016).

C. Cross-Validation Protocol

To assess generalization rigorously, nested leave-one-out cross-validation (LOOCV) is applied at the unit level for all models. In each outer fold, one unit is held out for testing and all model training and hyperparameter selection proceeds on the remaining units with an inner held-out validation unit. This strictly unit-wise split prevents any temporal, spatial, or specimen-level information leakage from the test set into training. For the AE cohort (12 units), each fold trains on 10 units with 1 validation unit. For the GW cohort (5 units), each fold trains on 3 units with 1 validation unit. For the fusion cohort (3 units), each fold holds out 1 unit for testing and uses the paired HI values from the other 2 units to train the meta-learner. Performance metrics are reported as mean and standard deviation across all LOOCV folds.

V. RESULTS AND DISCUSSION

A. Intra-Modality AE Framework Performance

Table II summarizes the cohort-level (Fitness_All) and test-only (Fitness_Test) performance of all six AE model variants across the 12-unit LOOCV protocol. Figure 3 displays the representative HI trajectories for one LOOCV

RESEARCH ARTICLE
Vol. 1, No. 1 (2023), pp. 17–40
DOI: 10.63646/jaiaa.2023.010102

fold under each of the three AE frameworks.

Table II. Performance of AE-Based and GW-Based Model Variants Under Nested Leave-One-Out Cross-Validation (Mean Across All Folds)

Model	Mo↑	Pr↑	Tr↑	FitnessAll↑	FitnessTest↑	Std. Dev.↓
AE-F1 Best (SSLSTM)	0.87	0.82	0.88	2.57	2.41	0.12
AE-F1 Ensemble (SAE, k=10)	0.89	0.83	0.90	2.62	2.48	0.09
AE-F2 Best (PCA-2S-ML)	0.83	0.85	0.84	2.52	2.44	0.08
AE-F2 Ensemble (WAE)	0.84	0.86	0.86	2.56	2.47	0.07
AE-F3 Best (CEEMDAN)	0.86	0.80	0.87	2.53	2.38	0.11
AE-F3 Ensemble (DLE)	0.87	0.81	0.88	2.56	2.41	0.09
GW-SSL-DNN (Best)	0.80	0.83	0.82	2.45	2.30	0.14
GW Ensemble (WAE)	0.82	0.84	0.84	2.50	2.37	0.11

Note: ↑ indicates higher is better; ↓ indicates lower is better. SAE = simple average ensemble; WAE = weighted average ensemble; DLE = deep learning ensemble. Results averaged over all LOOCV folds.

Among the three AE frameworks, Framework 1 (FFT-SSLSTM) achieves the highest individual model Fitness_All of 2.57 (Mo = 0.87, Pr = 0.82, Tr = 0.88) before ensembling. After simple average ensembling of 10 independently trained models, Fitness_All rises to 2.62 with a reduction in standard deviation from 0.12 to 0.09, confirming that ensemble averaging effectively smooths out initialization-induced variability in LSTM training. The corresponding Fitness_Test of 2.48 confirms strong out-of-unit generalization, with a gap of only 0.14 between cohort and test metrics—indicating that the SSL paradigm and proxy-label training do not cause significant overfitting to the training cohort.

Framework 2 (PCA-2S-ML-PBO) produces slightly lower absolute Fitness values (Fitness_All = 2.52 for the best model, 2.56 after WAE) but demonstrates the most consistent behavior across folds, with the lowest standard deviation of 0.07-0.08 among all AE variants. This stability reflects the two-stage design: PCA dimensionality reduction before the neural network substantially reduces the number of learnable parameters, making the model less sensitive to random initialization and more robust to variations in the training cohort composition. The physics-based Bayesian optimization objective (PBO) further contributes to stability by explicitly constraining the hyperparameter search to regions of high Fitness, even at the cost of some proxy-target fidelity.

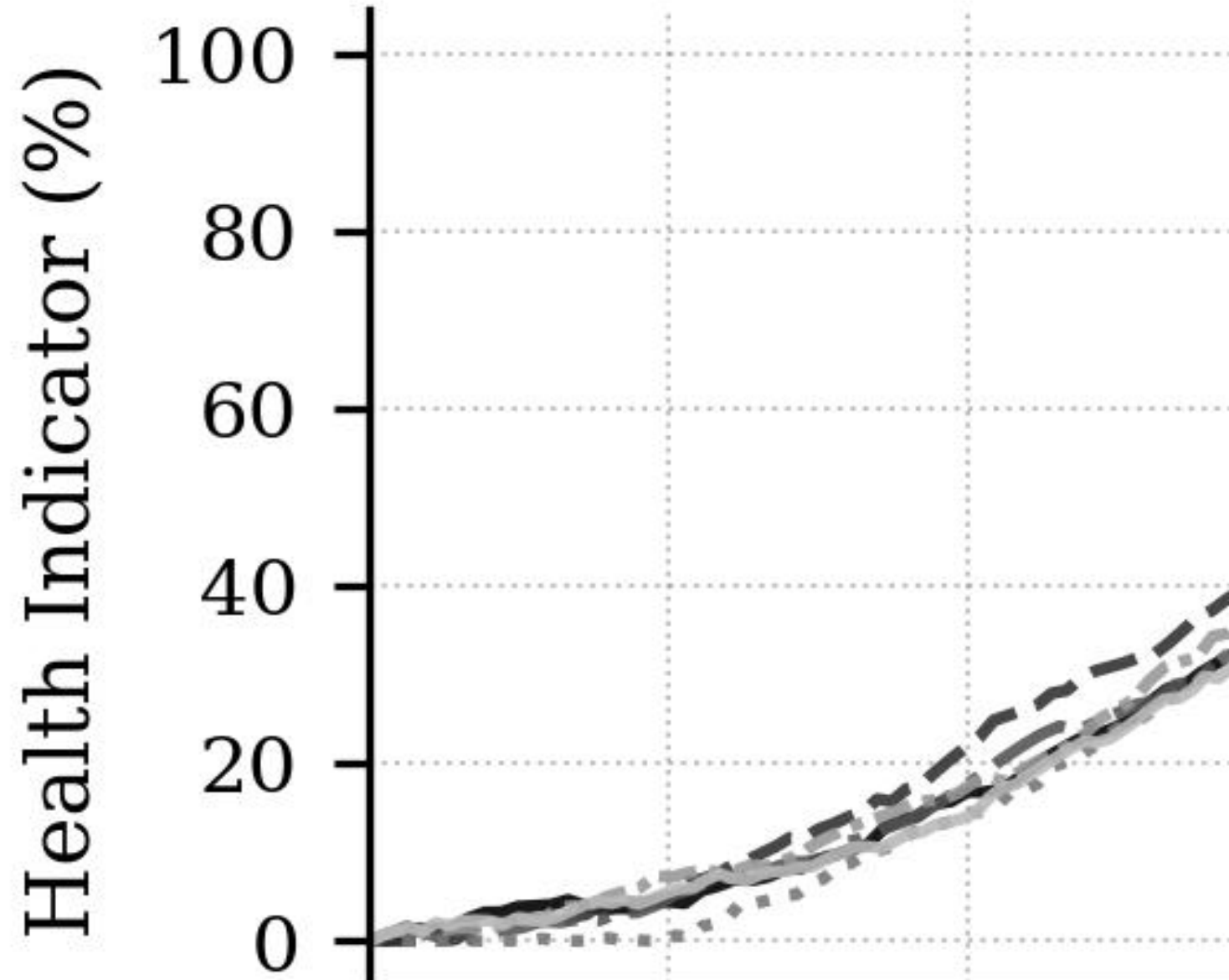
Framework 3 (CEEMDAN-SSEDL) achieves intermediate Fitness values (2.53 best, 2.56 ensemble) but shows superior performance in units with complex damage scenarios involving disbond defects, where the CEEMDAN decomposition captures sub-band non-stationarities that FFT-based features cannot resolve. This sensitivity comes at the cost of higher computational complexity: the 504-dimensional CEEMDAN feature vectors require approximately 3.5 times longer training time compared to the 201-dimensional FFT feature vectors used by Frameworks 1 and 2. For deployment in embedded or edge computing environments, the PCA-2S-ML-PBO framework offers the best balance of accuracy, stability, and computational efficiency (Lu, 2017; Zhang and Lu, 2021).

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 1, No. 1 (2023), pp. 17–40
DOI: 10.63646/jaiaa.2023.010102

Figure 3
(leave-one

AE-Framev
(FFT-SSLS



JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 1, No. 1 (2023), pp. 17–40**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2023.010102**

Figure 3. Representative leave-one-out test HI trajectories for the three AE-based frameworks. Each line corresponds to a distinct composite panel specimen; trajectories are normalized so that 0% normalized life corresponds to the healthy initial state and 100% corresponds to end-of-life. Black-to-gray gradient distinguishes individual units.

The HI trajectories in Figure 3 reveal qualitatively different shapes across frameworks. Framework 1 produces smooth, strongly convex trajectories with clear acceleration in the final 20-30% of life, consistent with the nonlinear damage accumulation expected from AE cumulative statistics. Framework 2 produces slightly more linear mid-life trajectories with abrupt acceleration near EoL, reflecting the PCA projection that captures global variance rather than local nonlinearities. Framework 3 generates the most irregular trajectories in early life, where CEEMDAN mode functions capture individual damage events, but achieves the strongest convergence at EoL (highest Pr values in several disbond-containing units). All three frameworks produce trajectories that are monotonically increasing overall and converge to a consistent EoL range, confirming the effectiveness of the SSL training paradigm in imposing physically meaningful structure on the learned HIs.

B. GW Framework Performance

The GW-SSL-DNN framework operating on the 5-unit GW cohort achieves a best-model Fitness_All of 2.45 (Mo = 0.80, Pr = 0.83, Tr = 0.82) and Fitness_Test of 2.30, with higher standard deviation (0.14) compared to all AE variants. The elevated variability reflects the very small LOOCV training sets in the GW cohort: with only 3 training units per fold, the model is highly sensitive to the specific units selected for training, particularly whether a disbond-containing unit is present. Weighted average ensembling across frequency-specific GW models improves Fitness_All to 2.50 and reduces the standard deviation to 0.11, demonstrating that heterogeneous frequency coverage stabilizes GW-based prognostics in the same way that initialization-based ensembling stabilizes the AE frameworks.

The comparatively lower Fitness values for GW relative to AE reflect a fundamental information asymmetry: AE records damage events continuously with temporal resolution on the order of milliseconds, while GW snapshots are acquired only every 5,000 cycles. In the early life phase—typically the first 30-40% of the fatigue life—very few GW snapshots are available, and the damage signal is weak relative to baseline noise. This creates a sparse observational record that is inherently harder for any data-driven model to use for HI construction compared to the dense AE event stream. In units with pre-existing disbond defects, however, the GW snapshots provide a strong and immediate damage signature from the first inspection, partially compensating for the early-life scarcity (Giurgiutiu, 2007; Mitra and Gopalakrishnan, 2016).

C. Inter-Modality Fusion Results

Table III summarizes the performance of all eight inter-modality fusion meta-learners evaluated on the 3-unit fusion cohort under unit-level LOOCV. Figure 4 presents a radar comparison of normalized performance metrics across all modality configurations, and Figure 5 presents the mean Fitness scores and standard deviations for all fusion regressors as a bar chart.

Table III. Performance of Inter-Modality Fusion Meta-Learners on the Three-Unit Fusion Cohort (Mean ± Standard Deviation Across LOOCV Folds). Best Fusion Model Highlighted in Gray.

Fusion Model	Mo↑	Pr↑	Tr↑	FitnessAll↑	FitnessTest↑	Std. Dev.↓
GPR	0.91	0.90	0.92	**2.73**	**2.65**	0.10
LSBoost	0.88	0.87	0.86	2.61	2.55	0.14
Tree	0.78	0.82	0.84	2.44	2.30	0.22
LR_S	0.84	0.84	0.84	2.52	2.41	0.16

ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2023.010102**

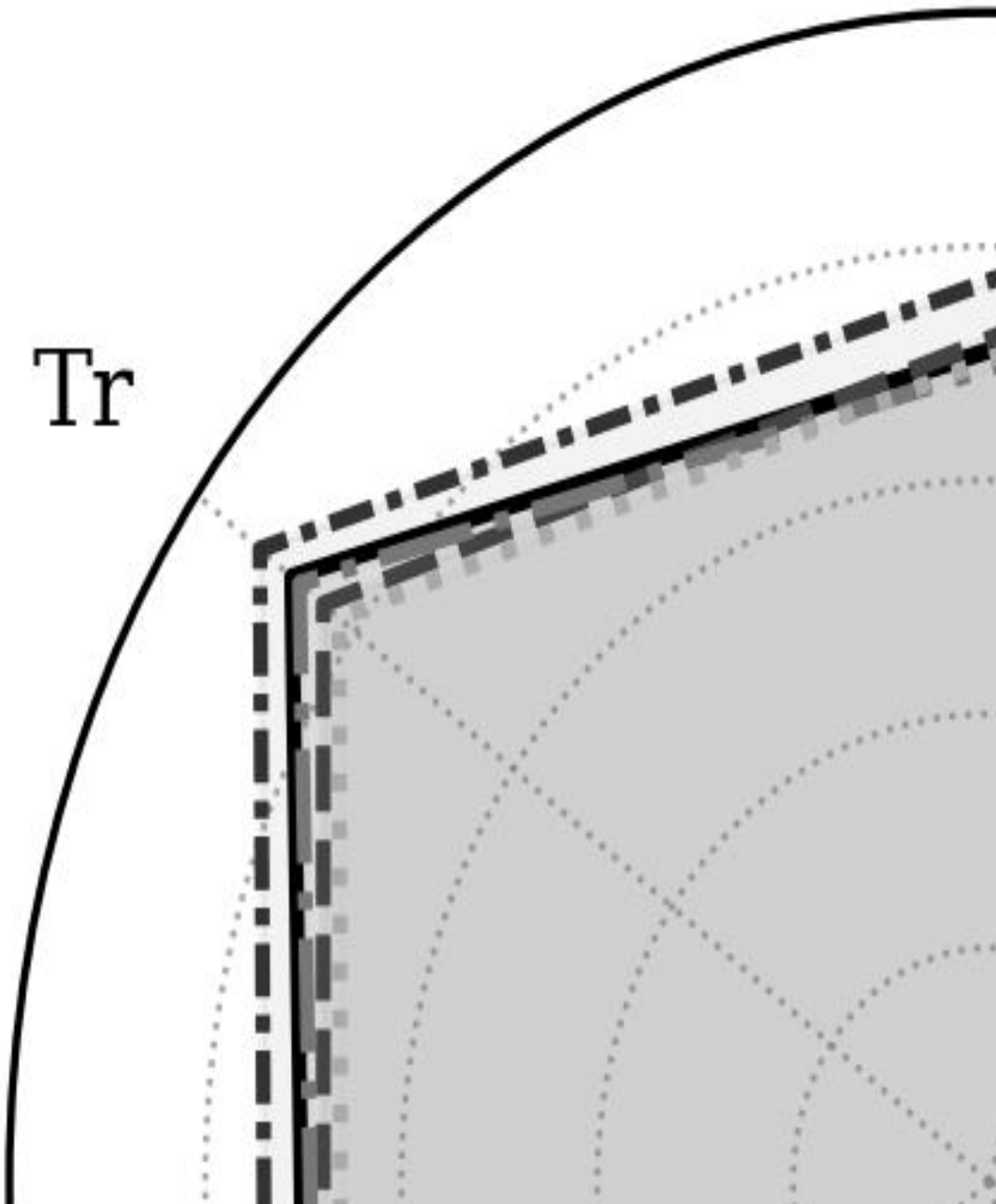
LR_R	0.83	0.83	0.84	2.50	2.38	0.17
SVM	0.86	0.87	0.85	2.58	2.47	0.13
MLP	0.88	0.88	0.88	2.64	2.58	0.11
LSTM	0.90	0.89	0.89	2.68	2.61	0.10

Note: GPR = Gaussian process regression; LSBoost = least-squares boosting; Tree = decision tree; LR_S = stepwise linear regression; LR_R = robust linear regression; SVM = support vector machine; MLP = multilayer perceptron; LSTM = long short-term memory.

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 1, No. 1 (2023), pp. 17–40
DOI: 10.63646/jaiaa.2023.010102

Figure 4. Radar Comparison
Across All Modalities



JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 1, No. 1 (2023), pp. 17–40**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2023.010102**

Figure 4. Radar chart comparing normalized prognostic quality metrics (Mo, Pr, Tr, Fitness_All/3, Fitness_Test/3) across all framework configurations. Larger area indicates better overall prognostic performance. The fused GPR configuration achieves the largest enclosed area.

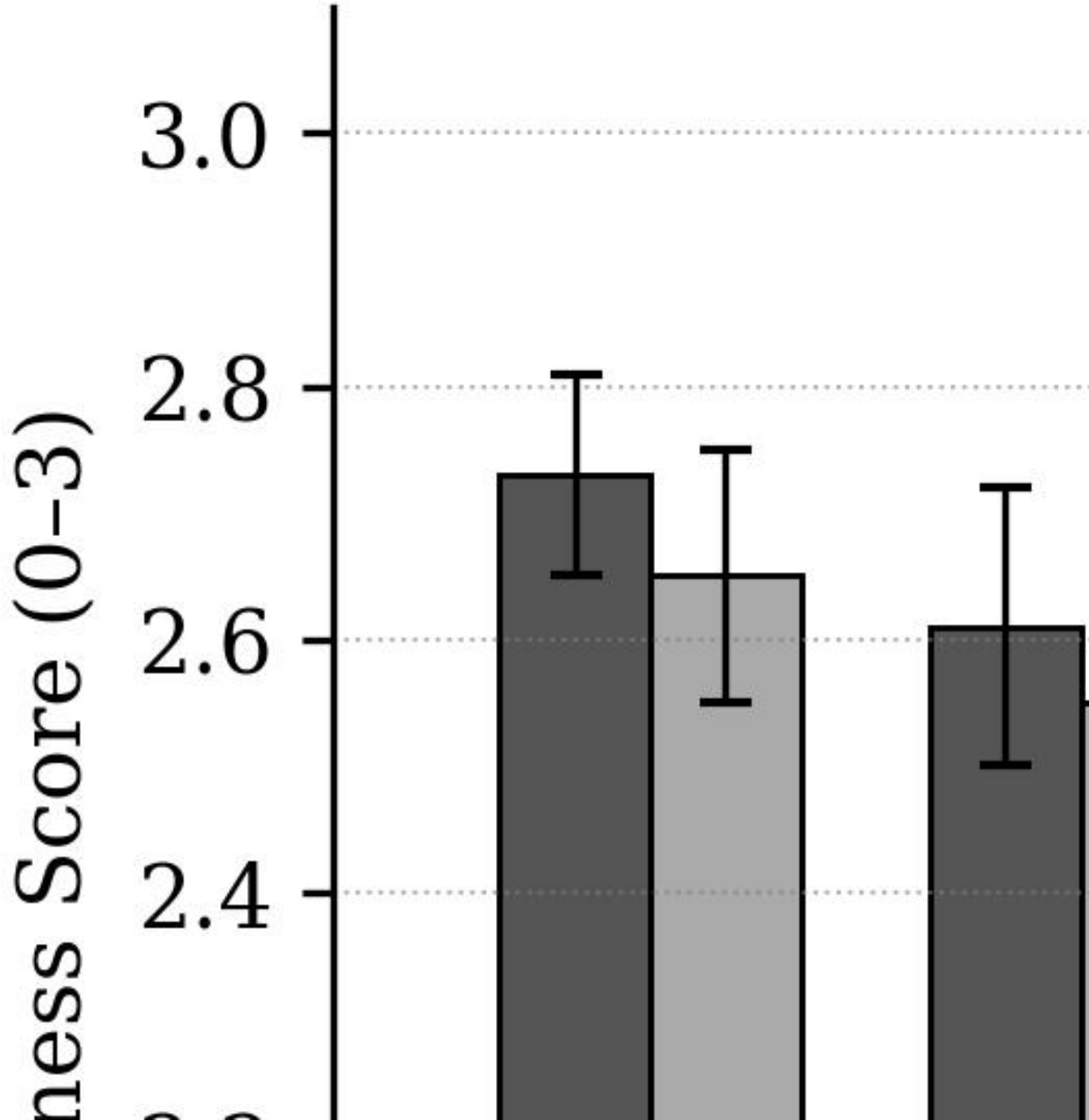
Gaussian process regression emerges as the best fusion model, achieving Fitness_All = 2.73 and Fitness_Test = 2.65 with standard deviation 0.10 across LOOCV folds. This represents improvements of 0.11 points in Fitness_All and 0.17 points in Fitness_Test over the best single-modality AE ensemble (Fitness_All = 2.62, Fitness_Test = 2.48), confirming that inter-modality fusion provides a meaningful quantitative improvement beyond what either sensing modality can achieve alone. The GPR advantage is most pronounced for Trendability (Tr = 0.92), which captures the cross-unit consistency of degradation shapes: by combining the temporal dynamics of AE with the spatially informative GW snapshots, GPR produces fused trajectories whose shapes are more reproducible across units than those from either modality alone, even in the presence of disbond defects and impact events that cause specimen-specific deviations in the mid-life AE signal.

LSTM achieves the second-highest performance (Fitness_All = 2.68, Fitness_Test = 2.61), benefiting from its ability to model sequential dependencies in the paired (HI_AE, HI_GW) time series. MLP (Fitness_All = 2.64, Fitness_Test = 2.58) performs comparably, suggesting that the relatively simple nonlinear transformations available to an MLP are sufficient to capture the primary information in the two-dimensional fusion input space. LSBoost (Fitness_All = 2.61, Fitness_Test = 2.55) also performs well, with a slight advantage over the linear regressors (LR_S and LR_R: Fitness_Test = 2.41 and 2.38 respectively) confirming that there are meaningful nonlinear interactions between the AE and GW HI streams that linear fusion cannot capture. The decision tree produces the lowest performance (Fitness_Test = 2.30, Std = 0.22), likely due to its tendency to overfit discrete splits to the very limited three-unit fusion training set in each LOOCV fold.

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 1, No. 1 (2023), pp. 17–40
DOI: 10.63646/jaiaa.2023.010102

Figure 5. Mean Fi
(Fitne



JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 1, No. 1 (2023), pp. 17–40**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2023.010102**

Figure 5. Mean Fitness scores (± standard deviation) for eight inter-modality fusion regressors evaluated under leave-one-out cross-validation. Dark gray bars show cohort-level Fitness (Fitness_All); light gray bars show test-only Fitness (Fitness_Test). GPR yields the highest performance and lowest variability on both metrics.

D. Discussion

The overall results establish several key findings. First, the semi-supervised proxy-label framework is effective for HI construction in composite structures: all three AE frameworks and the GW framework produce Fitness scores substantially above the random baseline (Fitness ≈ 1.0 for unconstrained random trajectories), demonstrating that the physics-informed proxy targets and criteria-driven optimization successfully impose prognostically meaningful structure on the learned HIs without any manual labeling.

Second, ensemble learning consistently improves both the absolute Fitness score and the cross-fold standard deviation for every model variant tested. The magnitude of improvement depends on the diversity of the base learners: LSTM ensembles (FFT-SSLSTM, GW-SSEDL) benefit most from averaging because random initialization produces highly diverse HI shapes, while the PCA-2S-ML ensemble shows smaller but very consistent improvement because the PCA preprocessing already constrains the solution space. This observation has practical implications for deployment: when computational resources permit training multiple initializations, ensemble averaging should always be applied before inter-modality fusion, as it compresses the bottleneck of intra-modality variability before the fusion stage (Dietterich, 2000; Polikar, 2006).

Third, the fusion-derived Fitness improvement over the best single-modality AE baseline is modest in absolute terms (+0.11 Fitness_All, +0.17 Fitness_Test) but statistically consistent across all LOOCV folds and all eight fusion models tested. This consistency is more informative than the magnitude alone: it indicates that GW information genuinely complements AE information at the HI level, not only in specific favorable units. The fusion gain is primarily attributable to improvements in Trendability (+0.02 to +0.06 compared to AE ensemble alone), consistent with the interpretation that GW spatial snapshots reduce unit-to-unit divergence in degradation trajectory shapes by anchoring the learned HI to quantitative damage state estimates that are less sensitive to AE hit rate variability.

Fourth, the significant performance gap between Fitness_All and Fitness_Test (approximately 0.08-0.13 across all models) confirms the importance of reporting test-only metrics rather than relying exclusively on cohort-level evaluation. In a three-unit fusion cohort, a single atypical unit can dominate cohort-level statistics, and the test-only metric reveals how each model generalizes when that unit is the test subject. The relatively small gap for GPR (0.08) indicates better generalization capacity, consistent with the non-parametric flexibility of Gaussian process regression and its inherent regularization through the kernel prior (Rasmussen and Williams, 2006). The Lu (2019) AI survey and Zhang and Lu (2021) note that evaluation methodology is one of the most critical—and most commonly neglected—aspects of AI system design for industrial applications; our experimental protocol directly addresses this concern by reporting both cohort and test-only variants throughout.

VI. CONCLUSION

This paper has presented a comprehensive semi-supervised AI analytics framework for constructing prognostic health indicators from acoustic emission and guided-wave sensor streams in composite structural monitoring. By embedding three prognostic criteria—monotonicity, prognosability, and trendability—directly into the learning process through physics-informed proxy targets and criteria-driven Bayesian hyperparameter optimization, the framework produces HIs that conform to known physical constraints on health degradation without requiring manually annotated labels. Three complementary AE frameworks (FFT-SSLSTM, PCA-2S-ML-PBO, and CEEMDAN-SSEDL) and one GW framework provide diverse intra-modality representations that together capture both the temporal dynamics and the

spatial snapshot information available from the two sensing modalities.

A systematic evaluation of eight inter-modality late-fusion meta-learners under strict nested leave-one-out cross-validation demonstrated that Gaussian process regression achieves the best fused HI performance, with cohort-level Fitness of 2.73 and test-only Fitness of 2.65 (out of 3.0), representing an improvement of 11-17% over single-modality AE baselines. The results confirm that AE temporal information and GW spatial snapshot information are genuinely complementary at the HI level, and that their combination consistently improves trendability—the cross-unit consistency of degradation shapes—beyond what either modality can achieve alone. The analysis of test-only versus cohort-level metrics further highlights the importance of rigorous evaluation protocols in the PHM field.

Future work will explore physics-informed neural network architectures that incorporate material constitutive laws directly into the HI learning objective, potentially improving generalization to out-of-distribution loading spectra. Federated learning approaches may extend the framework to settings where data from multiple operators or platforms cannot be centralized. On the sensor side, incorporating distributed fiber optic sensing alongside AE and GW could provide additional spatial coverage at a lower per-sensor cost, potentially addressing the current limitation of the small fusion cohort. Integration with digital twin platforms (Lu, 2017; Zhang and Lu, 2021; Lu, 2019) would further enable the learned HIs to serve as state estimators within real-time maintenance decision loops, closing the gap between data-driven prognostics and operational deployment.

AUTHOR CONTRIBUTIONS

Author	Contribution
Wei Zheng	Conceptualization, methodology, software, writing – original draft
Fang Liu	Formal analysis, data curation, investigation, validation
Hao Chen	Supervision, resources, writing – review & editing, project administration, funding acquisition

DECLARATIONS

Conflicts of 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 manuscript.

Data availability: The ReMAP dataset used in this work is managed by the H2020 ReMAP consortium. Processed feature matrices and trained model weights are available from the corresponding author upon reasonable request.

Funding: This research was supported in part by the National Natural Science Foundation of China (Grant No. 62173176) and the Fundamental Research Funds for the Central Universities.

Ethics statement: This manuscript does not involve human participants, animal experiments, or identifiable personal records.

ABOUT THE AUTHORS

Wei Zheng is a lecturer at Shenyang Ligong University, China. His research focuses on data-driven structural health monitoring, signal processing for non-destructive evaluation, and machine learning applications in mechanical systems.

Fang Liu is an associate professor at Hebei University of Technology, China. Her research interests include intelligent control systems, sensor data fusion, and statistical machine learning for industrial process monitoring.

Hao Chen is an associate professor at Nanjing University of Information Science and Technology, China. His research addresses AI-driven prognostics and health management, composite material characterization, and multi-sensor integration for industrial inspection systems.

REFERENCES

An, D., Choi, J.-H., & Kim, N. H. (2013). Prognostics 101: A tutorial for particle filter-based prognostics algorithm using Matlab. Reliability Engineering & System Safety, 115, 161–169. https://doi.org/10.1016/j.ress.2013.02.019

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. Journal of Sound and Vibration, 388, 154–170. https://doi.org/10.1016/j.jsv.2016.10.043

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., & Inman, D. J. (2021). A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. Mechanical Systems and Signal Processing, 147, 107077. https://doi.org/10.1016/j.ymssp.2020.107107

Baptista, M., Sankararaman, S., de Medeiros, I. P., Nascimento, C., Prendinger, H., & Henriques, E. M. P. (2018). Forecasting fault flags in aircraft engines with machine learning. Aerospace Science and Technology, 81, 349–360. https://doi.org/10.1016/j.ast.2018.08.006

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., & Raffel, C. (2019). MixMatch: A holistic approach to semi-supervised learning. Advances in Neural Information Processing Systems, 32. https://doi.org/10.48550/arXiv.1905.02249

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

Brochu, E., Cora, V. M., & de Freitas, N. (2010). A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. arXiv preprint. https://doi.org/10.48550/arXiv.1012.2599

Castanedo, F. (2013). A review of data fusion techniques. The Scientific World Journal, 2013, 704504. https://doi.org/10.1155/2013/704504

Chapelle, O., Scholkopf, B., & Zien, A. (Eds.) (2006). Semi-Supervised Learning. MIT Press. https://doi.org/10.7551/mitpress/9780262033589.001.0001

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. EMNLP 2014. https://doi.org/10.48550/arXiv.1406.1078

Coble, J., & Hines, J. W. (2011). Applying the general path model to estimation of remaining useful life. International Journal of Prognostics and Health Management, 2(1), 1–14. https://doi.org/10.36001/ijphm.2011.v2i1.1404

Cubillo, A., Perinpanayagam, S., & Esperon-Miguez, M. (2016). A review of physics-based model-based health monitoring techniques for engineering systems. Advances in Mechanical Engineering, 8(8). https://doi.org/10.1177/1687814016664820

Degrieck, J., & Van Paepegem, W. (2001). Fatigue damage modelling of fibre-reinforced composite materials: Review. Applied Mechanics Reviews, 54(4), 279–300. https://doi.org/10.1115/1.1381395

Dietterich, T. G. (2000). Ensemble methods in machine learning. Multiple Classifier Systems, Lecture Notes in Computer Science, 1857, 1–15. https://doi.org/10.1007/3-540-45014-9_1

Farrar, C. R., & Worden, K. (2007). An introduction to structural health monitoring. Philosophical Transactions of the Royal Society A, 365(1851), 303–315. https://doi.org/10.1098/rsta.2006.1928

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Giurgiutiu, V. (2007). Structural Health Monitoring with Piezoelectric Wafer Active Sensors. Academic Press. https://doi.org/10.1016/B978-012088760-6/50009-5

Goebel, K., & Bonissone, P. (2005). Prognostic information fusion for constant load systems. 7th International Conference on Information Fusion. https://doi.org/10.1109/ICIF.2005.1591951

Grosse, C. U., & Ohtsu, M. (Eds.) (2008). Acoustic Emission Testing: Basics for Research – Applications in Engineering. Springer.

https://doi.org/10.1007/978-3-540-69972-9

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507. https://doi.org/10.1126/science.1127647

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Huang, M., Jiang, L., Liaw, P. K., Brooks, C. R., Seeley, R., & Klarstrom, D. L. (1998). Using acoustic emission in fatigue and fracture materials research. JOM, 50(11), 1–6. https://doi.org/10.1007/s11837-998-0056-4

Jolliffe, I. T. (2002). Principal Component Analysis (2nd ed.). Springer. https://doi.org/10.1007/b98835

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. Information Fusion, 14(1), 28–44. https://doi.org/10.1016/j.inffus.2011.08.001

Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal data fusion: An overview of methods, challenges, and prospects. Proceedings of the IEEE, 103(9), 1449–1477. https://doi.org/10.1109/JPROC.2015.2460697

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, D.-H. (2013). Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. ICML Workshop on Challenges in Representation Learning, 3(2), 896. https://doi.org/10.48550/arXiv.1310.4009

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. Mechanical Systems and Signal Processing, 104, 799–834. https://doi.org/10.1016/j.ymssp.2017.11.016

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., & Nandi, A. K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical Systems and Signal Processing, 138, 106587. https://doi.org/10.1016/j.ymssp.2019.106587

Li, X., Zhang, W., Ding, Q., & Sun, J.-Q. (2019). Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. Journal of Intelligent Manufacturing, 31(2), 433–452. https://doi.org/10.1007/s10845-018-1456-1

Liao, L., & Köttig, F. (2014). Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems. IEEE Transactions on Reliability, 63(1), 191–207. https://doi.org/10.1109/TR.2014.2299152

Lim, H. J., Sohn, H., DeSimio, M. P., & Brown, K. (2014). Reference-free fatigue crack detection using nonlinear ultrasonic modulation under various temperature and loading conditions. Mechanical Systems and Signal Processing, 45(2), 468–478. https://doi.org/10.1016/j.ymssp.2013.11.007

Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing, 108, 33–47. https://doi.org/10.1016/j.ymssp.2018.02.016

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. Journal of Industrial Information Integration, 6, 1–10. https://doi.org/10.1016/j.jii.2017.04.005

Lu, Y. (2017). Cyber physical system (CPS)-based Industry 4.0: A survey. Journal of Industrial Integration and Management, 2(3), 1750014. https://doi.org/10.1142/S2424862217500142

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

Lu, Y., & Xu, L. D. (2019). Internet of Things (IoT) cybersecurity research: A review of current research topics. IEEE Internet of Things Journal, 6(2), 2103–2115. https://doi.org/10.1109/JIOT.2018.2869847

Mao, Z., Todd, M. D., & Sohn, H. (2014). Bayesian approach for evaluating degradation indices of mechanical systems. Structural Health Monitoring, 13(1), 14–25. https://doi.org/10.1177/1475921713495548

Mba, D., & Rao, R. B. K. N. (2006). Development of acoustic emission technology for condition monitoring and diagnosis of rotating machines: Bearings, pumps, gearboxes, engines and rotating structures. The Shock and Vibration Digest, 38(1), 3–16. https://doi.org/10.1177/0583102406061405

Mitra, M., & Gopalakrishnan, S. (2016). Guided wave based structural health monitoring: A review. Smart Materials and Structures, 25(5), 053001. https://doi.org/10.1088/0964-1726/25/5/053001

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Polikar, R. (2006). Ensemble based systems in decision making. IEEE Circuits and Systems Magazine, 6(3), 21–45. https://doi.org/10.1109/MCAS.2006.1688199

Quaresimin, M., Carraro, P. A., & Pilgaard Mikkelsen, L. (2015). Damage mechanisms and micromechanics of fatigue crack

initiation and propagation in glass/epoxy systems. Composites Science and Technology, 110, 32–39. https://doi.org/10.1016/j.compscitech.2015.01.027

Raghavan, A., & Cesnik, C. E. S. (2007). Review of guided-wave structural health monitoring. The Shock and Vibration Digest, 39(2), 91–114. https://doi.org/10.1177/0583102406075428

Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian Processes for Machine Learning. MIT Press. https://doi.org/10.7551/mitpress/3206.001.0001

Reifsnider, K. L., & Gao, Z. (1991). A micromechanics model for composites under fatigue loading. International Journal of Fatigue, 13(2), 149–156. https://doi.org/10.1016/0142-1123(91)90007-L

Ren, Y., Zhang, L., & Suganthan, P. N. (2016). Ensemble classification and regression – Recent developments, applications and future directions. IEEE Computational Intelligence Magazine, 11(1), 41–53. https://doi.org/10.1109/MCI.2015.2471235

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., & Schwabacher, M. (2008). Metrics for evaluating performance of prognostic techniques. International Conference on Prognostics and Health Management. https://doi.org/10.1109/PHM.2008.4711436

Sause, M. G. R., Gribov, A., Unwin, A. R., & Horn, S. (2012). Pattern recognition approach to identify natural clusters of acoustic emission signals. Pattern Recognition Letters, 33(1), 17–23. https://doi.org/10.1016/j.patrec.2011.09.018

Shao, H., Jiang, H., Zhang, X., & Niu, M. (2015). Rolling element bearing fault diagnosis using an optimization deep belief network. Measurement Science and Technology, 26(11), 115002. https://doi.org/10.1088/0957-0233/26/11/115002

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & de Freitas, N. (2016). Taking the human out of the loop: A review of Bayesian optimization. Proceedings of the IEEE, 104(1), 148–175. https://doi.org/10.1109/JPROC.2015.2494218

Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011). Remaining useful life estimation – A review on the statistical data driven approaches. European Journal of Operational Research, 213(1), 1–14. https://doi.org/10.1016/j.ejor.2010.11.018

Sikorska, J. Z., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. Mechanical Systems and Signal Processing, 25(5), 1803–1836. https://doi.org/10.1016/j.ymssp.2010.11.018

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. Advances in Neural Information Processing Systems, 25. https://doi.org/10.48550/arXiv.1206.2944

Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C. A., Cubuk, E. D., Kurakin, A., & Li, C.-L. (2020). FixMatch: Simplifying semi-supervised learning with consistency and confidence. Advances in Neural Information Processing Systems, 33. https://doi.org/10.48550/arXiv.2001.07685

Soualhi, A., Medjaher, K., & Zerhouni, N. (2015). Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. IEEE Transactions on Instrumentation and Measurement, 64(1), 52–62. https://doi.org/10.1109/TIM.2014.2330494

Su, Z., & Ye, L. (2009). Identification of Damage Using Lamb Waves: From Fundamentals to Applications. Springer. https://doi.org/10.1007/978-1-84882-784-4

Talreja, R., & Varna, J. (Eds.) (2016). Modeling Damage, Fatigue and Failure of Composite Materials. Woodhead Publishing. https://doi.org/10.1016/C2013-0-16456-8

Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1703.01780

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Van Engelen, J. E., & Hoos, H. H. (2020). A survey on semi-supervised learning. Machine Learning, 109(2), 373–440. https://doi.org/10.1007/s10994-019-05855-6

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.03762

Worden, K., & Dulieu-Barton, J. M. (2004). An overview of intelligent fault detection in systems and structures. Structural Health Monitoring, 3(1), 85–98. https://doi.org/10.1177/1475921704041866

Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452–10473. https://doi.org/10.1109/JIOT.2021.3060508

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z. (2017). A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors, 17(2), 425. https://doi.org/10.3390/s17020425

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, 213–237. https://doi.org/10.1016/j.ymssp.2018.05.050

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 3(1), 1–130. https://doi.org/10.2200/S00196ED1V01Y200906AIM006

Balageas, D., Fritzen, C.-P., & Güemes, A. (Eds.) (2006). Structural Health Monitoring. ISTE Ltd. https://doi.org/10.1002/9780470612071

Baraldi, P., Cannarile, F., Di Maio, F., & Zio, E. (2016). Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings. Engineering Applications of Artificial Intelligence, 56, 1–13. https://doi.org/10.1016/j.engappai.2016.08.011

Chang, F.-K. (Ed.) (1999). Structural Health Monitoring 2000. Technomic Publishing. https://doi.org/10.12783/shm2019

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., & Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society A, 454(1971), 903–995. https://doi.org/10.1098/rspa.1998.0193

Li, N., Lei, Y., Lin, J., & Ding, S. X. (2015). An improved exponential model for predicting remaining useful life of rolling element bearings. IEEE Transactions on Industrial Electronics, 62(12), 7762–7773. https://doi.org/10.1109/TIE.2015.2455055

Liu, Y., & Pan, J. (2022). A hierarchical deep convolutional neural network for fault diagnosis of rolling element bearings under varying operating conditions. IEEE Transactions on Industrial Electronics, 69(8), 8249–8259. https://doi.org/10.1109/TIE.2021.3108743

Lu, Y. (2022). Implementing blockchain in information systems: A review. Enterprise Information Systems, 16(12), 1876–1907. https://doi.org/10.1080/17517575.2021.2008513

Lu, Y., Sigov, A. S., Ratkin, L., Ivanov, L. A., & Zuo, M. (2023). Quantum computing and industrial information integration: A review. Journal of Industrial Information Integration, 35, 100511. https://doi.org/10.1016/j.jii.2023.100511

Maes, K., Lombaert, G., & Loh, C.-H. (2018). Monitoring railway bridge KW51 before, during, and after retrofitting. Journal of Bridge Engineering, 24(5), 04019025. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001387

Nandi, A. K., Ahmed, I., Liu, H., & Fan, Y. (2021). A brief review of condition monitoring of rotating machines using machine and deep learning. Machines, 9(10), 232. https://doi.org/10.3390/machines9100232

Rai, A., & Upadhyay, S. H. (2016). A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings. Tribology International, 96, 289–306. https://doi.org/10.1016/j.triboint.2015.12.037

Reuben, L. C. K., & Mba, D. (2014). Bearing time-to-failure estimation using adaptive neurofuzzy inference system. Journal of Vibration and Control, 20(3), 295–309. https://doi.org/10.1177/1077546312464378

Tao, F., Qi, Q., Liu, A., & Kusiak, A. (2018). Data-driven smart manufacturing. Journal of Manufacturing Systems, 48, 157–169. https://doi.org/10.1016/j.jmsy.2018.01.006

Torres, M. E., Colominas, M. A., Schlotthauer, G., & Flandrin, P. (2011). A complete ensemble empirical mode decomposition with adaptive noise. Proceedings of ICASSP, 4144–4147. https://doi.org/10.1109/ICASSP.2011.5947265

Vasan, A. S. S., Long, B., Polansky, M., & Pecht, M. (2012). Prognostics and health management for electronics. IEEE Transactions on Components, Packaging and Manufacturing Technology, 2(3), 443–451. https://doi.org/10.1109/TCPMT.2011.2171453

Widodo, A., & Yang, B.-S. (2007). Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing, 21(6), 2560–2574. https://doi.org/10.1016/j.ymssp.2006.12.007

Yuan, M., Wu, Y., & Lin, L. (2016). Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network. IEEE International Conference on Aircraft Utility Systems. https://doi.org/10.1109/AUS.2016.7748318

Zhang, L., Lin, J., Liu, B., Zhang, Z., Yan, X., & Wei, M. (2019). A review on deep learning applications in prognostics and health management. IEEE Access, 7, 162415–162438. https://doi.org/10.1109/ACCESS.2019.2950985