POLITECNICO DI TORINO
Repository ISTITUZIONALE

Element-level finite-element-informed machine-learning surrogates for failure-index prediction in a composite wing box with intralaminar damage

Original
Element-level finite-element-informed machine-learning surrogates for failure-index prediction in a composite wing box with intralaminar damage / Magliacano, D., Letizia, A., Di Palo, S., Alecci, M., Bazzani, M.. - In: AEROSPACE SCIENCE AND TECHNOLOGY. - ISSN 1270-9638. - ELETTRONICO. - 177 D:(2026), pp. 1-19. [10.1016/j.ast.2026.112878]

Availability:
This version is available at: 11583/3012149 since: 2026-06-17T07:40:34Z

Publisher:
Elsevier

Published
DOI:10.1016/j.ast.2026.112878

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

16 July 2026

Aerospace Science and Technology 177 (2026) 112878

Contents lists available at ScienceDirect

Aerospace Science and Technology

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

Element-level finite-element-informed machine-learning surrogates for failure-index prediction in a composite wing box with intralaminar damage

Dario Magliacano [a,*], Annalisa Letizia [b], Sara Di Palo [b], Marika Alecci [b], Marco Bazzani [b]

[a] Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, Turin, 10129, TO, Italy
[b] Teoresi S.p.A., Via Perugia 24, Turin, 10152, TO, Italy

ARTICLE INFO

Communicated by Dr Mehdi Ghoreyshi

Keywords:
Machine learning
Finite-element surrogate
Failure index
Composite wing box
Intralaminar damage
Structural health monitoring

ABSTRACT

This study develops and numerically assesses finite-element-informed machine-learning surrogates for predicting five classical failure indices in a composite wing-box substructure containing prescribed intralaminar stiffness degradation. A damage-oriented database is generated by varying damage position, extent, and intensity over rib, skin, and spar regions of an equivalent-orthotropic finite-element model. Random Forest, XGBoost, LightGBM, and a multilayer perceptron are trained and evaluated using a common data-partitioning and cross-validation framework, with separate attention to the upper end of the failure-index range. The tree-based ensembles consistently outperform the neural-network benchmark; the Max Stress index is the most accurately reproduced, whereas the Christensen index and spar-dominated cases remain more difficult, particularly for failure-index values above 0.7. The results establish the feasibility of rapid numerical screening within the adopted model assumptions.

1. Introduction

Composite wing boxes provide high specific stiffness and strength, but their structural response can be strongly affected by spatially distributed intralaminar degradation. Repeated finite-element analyses can quantify the associated changes in local stresses and failure indices, although their computational cost limits their use for rapid screening of numerous damage configurations. The present study therefore investigates machine-learning surrogates that reproduce the failure-index response of a prescribed finite-element model from damage and geometric descriptors. In this context, the predicted failure indices are treated as static structural-response indicators rather than estimates of remaining fatigue life or time to failure.

1.1. Background and motivation

Traditional approaches to predicting structural health in aircraft structures often rely on computationally expensive models and established damage mechanics principles. In this field, Chehade and Younes [1] reviewed various software and calculation tools used for structural reliability and health prediction. One prevalent method involves strength-to-failure analysis. This approach calculates the load required to cause complete failure in an undamaged wing box. An estimate of structural health can be made by comparing this value to the expected maximum loads during operation. However, this method has limitations. It doesn't account for gradual degradation and the growth of potential damage over time, which can lead to an underestimation of structural health. Fanteria and Panettieri [2] presented a constitutive model for the nonlinear shear behavior of composite laminates, developed using a user-defined Fortran routine that could be used in the Abaqus nonlinear finite element (FE) code. Carrera Unified Formulation (CUF) offers a consistent framework for approximating a wide range of structural configurations, such as beams, plates, and shells, by formulating the displacement field as a series of expansion functions rather than in traditional layer-wise or equivalent single-layer formulations. With this strategy, high-order accuracy [3] and the capability to simulate complex strain and displacement fields can be achieved using a greater number of expansion functions [4]. Moreover, CUF can also handle structures of various geometries and can simulate non-linear behavior, which also includes geometric non-linearities [5,6].

Data-driven structural health monitoring studies have predominantly focused on detecting, locating, or classifying damage from vibration, guided-wave, or other sensor-derived information [7–11]. Complementary FE-informed approaches employ numerical simulations to generate labeled data for surrogate modeling, thereby enabling rapid response or failure prediction once the surrogate has been trained [12,13]. Alaimo et al. [14] investigated an artificial neural network (ANN) that can detect and localize damage without prior knowledge of its

* Corresponding author.
E-mail address: darius6591@hotmail.it;dario.magliacano@unina.it;dario.magliacano@polito.it (D. Magliacano).

https://doi.org/10.1016/j.ast.2026.112878
Received 9 December 2025; Received in revised form 5 June 2026; Accepted 10 June 2026
Available online 17 June 2026
1270-9638/© 2026 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

characteristics to serve as a real-time data processor for SHM systems. Casaburo et al. [15] developed and validated a machine learning (ML)-based tool for the design of global vibroacoustic treatments based on foam cores with embedded periodic patterns, which allow passive control of acoustic paths in layered concepts. Azad and Kim [16] used hybrid deep convolutional networks to propose a robust autonomous damage-diagnosis method for laminated composite structures. Wang et al. [12] predicted the failure factors of composite pressure vessels using a combined FEM simulation and machine learning approach. Early applications have already demonstrated the feasibility of combining long-term structural health monitoring systems with neural-network-based damage identification in large-scale real-world structures, such as bridges [17,18].

These approaches are suitable for controlled numerical investigations, whereas transfer to physical structures remains a separate task that requires experimental calibration and validation. The use of a wing-box geometry is relevant because it introduces interacting ribs, skins, and spars, as well as non-uniform stress distributions not represented by coupon-level or simple-panel studies. Nevertheless, the present model is deliberately simplified to isolate the relationship between prescribed intralaminar stiffness degradation and the resulting failure indices.

1.2. Related work and contribution

The previously published wing-box benchmark [19] varied orthotropic material properties and compared Random Forest and Support Vector Regression surrogates using failure indices evaluated at the most-stressed element. The present study instead fixes a single equivalent orthotropic material to isolate damage-related effects and develops an element-level, damage-oriented database spanning rib, skin, and spar regions. Random Forest, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and a multilayer perceptron are evaluated using a common validation procedure, and model behavior is examined separately in the upper range of the failure index.

Accordingly, the contribution does not consist of a new machine-learning algorithm. It consists of an application-specific element-level database, a comparative model-selection study, and a safety-oriented analysis of prediction errors close to the critical failure-index range. The study is a numerical proof of concept; it does not reproduce stacking-sequence effects, delamination, the evolution of progressive physical damage, or experimentally measured structural responses.

Section 2 describes the numerical model, prescribed damage representation, database generation, and learning framework. Section 3 presents the comparative results, including the dedicated upper-range error analysis. Section 4 summarizes the principal findings, limitations, and required developments.

2. Methodological framework

The methodology comprises two connected stages. First, a consistently discretized finite-element model is used to generate labeled structural-response data for prescribed damage configurations. Second, several regression models are trained to reproduce the resulting failure indices from damage and geometric descriptors. Because all target values originate from the finite-element model, the assessment quantifies surrogate fidelity with respect to that model. The key to developing effective AI models lies in the relevance and quality of the data used to train them. In this section, the methods for data acquisition and preprocessing applied in this study are described.

2.1. Geometry and material properties

The initial data-acquisition stage focuses on collecting information on the geometry and material properties of the composite wing box, which serves as the reference component. This data serves as a crucial foundation for the AI model, enabling it to understand the specific characteristics of the structure under evaluation. The reference component is a two-bay wing-box substructure extracted from a freely available CAD representation of the Cirrus SR22 aircraft, shown in Fig. 1. The geometry is used solely as a representative aerospace wing-box configuration and is not intended to reproduce the proprietary or certified structural definition of the actual aircraft.

A single equivalent orthotropic continuum is adopted to isolate the mapping between prescribed intralaminar stiffness degradation and the resulting failure-index response. This representation deliberately excludes stacking-sequence-dependent behavior, ply-level interlaminar stresses, and delamination initiation or propagation. Reducing the number of laminate-specific variables allows the influence of damage location, extent, and intensity to be examined under controlled conditions. Consequently, the results apply to the adopted equivalent-material model and should not be interpreted as a ply-resolved assessment of the actual aircraft structure. The mechanical properties and ultimate strengths used for this equivalent representation are reported in Tables 1 and 2, respectively, while the chosen material density is equal to $1600\,\mathrm{kg/m^3}$.

Table 1
Equivalent orthotropic mechanical properties of the analyzed carbon fiber unidirectional composite material [20].

Young Modulus [Pa]		
E_{11}	E_{22}	E_{33}
1.35E+11	1.00E+10	1.00E+10
Shear Modulus [Pa]		
G_{12}	G_{13}	G_{23}
5.00E+09	5.00E+09	3.60E+09
Poisson Ratio		
ν_{12}	ν_{13}	ν_{23}
0.3	0.3	0.39

Table 2
Equivalent orthotropic ultimate strengths of the analyzed carbon fiber unidirectional composite material.

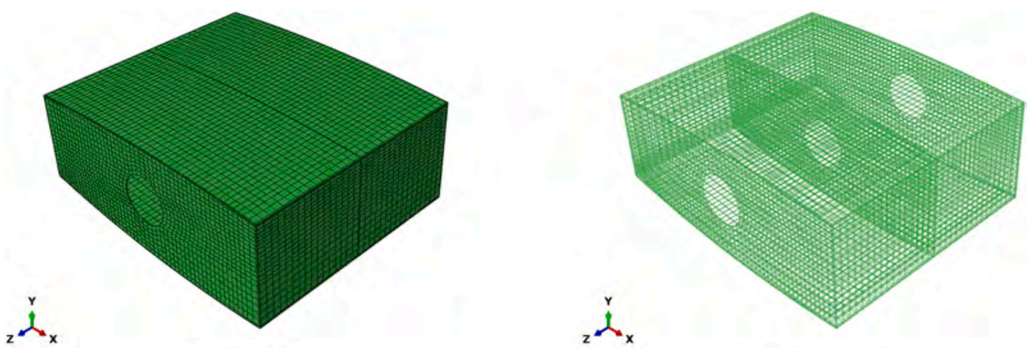
Ultimate Strength [Pa]		
S_{11+}	S_{11-}	S_{22+}
1.50E+09	1.20E+09	5.00E+07
Ultimate Strength [Pa]		
S_{22-}	S_{12}	S_{23}
2.50E+08	7.00E+07	5.04E+07

- S_{11+} is the value of stress component σ_{11} at longitudinal tensile failure;
- S_{11-} is the value of stress component σ_{11} at compressive tensile failure;
- S_{22+} is the value of stress component σ_{22} at longitudinal tensile failure;
- S_{22-} is the value of stress component σ_{22} at compressive tensile failure;
- S_{12} is the absolute value of stress component σ_{12} at longitudinal shear failure;
- S_{23} is the absolute value of stress component σ_{23} at transverse shear failure.

In this regard, the ANN model can establish relationships between the physical characteristics of the damaged wing box and its remaining structural health by incorporating comprehensive geometric and material property data.

(a) Front. (b) Rear.

Fig. 1. Cirrus SR22 aircraft CAD isometric views.



(a) Solid. (b) Transparent.

Fig. 2. Reference wing box isometric views.

2.2. Numerical model characteristics, constraints and loads

The extracted component from the reference aircraft is constituted by two half-wing bays, meaning a portion of the half-wing is delimited by three ribs. The thicknesses of the main elements of the reference wing box, including ribs, skins, and spars, are 2.5 mm, 2.0 mm, 3.0 mm, respectively. Fig. 2 shows the analyzed wing box's solid and transparent isometric views.

By numbering the ribs from 1 to 3, starting from the closest one, boundary conditions are constituted by a fixed constraint applied on the external surface of Rib 3 (Fig. 3a), simulating the wing-fuselage junction. Distributed static loads are applied on the external surface of Rib 1 (Fig. 3b): they are meant to represent the most critical operating condition, which is identified by the so-called maximum maneuver point (located in the intersection between the maximum load and stall curves) in the flight envelope plot. More specifically, these load components are estimated to be: $F_x = 0\,\mathrm{N}$, $F_y = -1774.88\,\mathrm{N}$, $F_z = 21\,124.1\,\mathrm{N}$, $M_x = -43\,640.3\,\mathrm{N\,m}$, $M_y = 5352.03\,\mathrm{N\,m}$, $M_z = 389.231\,\mathrm{N\,m}$; for failure indices (FIs) estimation, these values are increased by a safety factor equal to 1.5. Furthermore, a gravity load equal to $-9.81\,\frac{\mathrm{m}}{\mathrm{s}^2}$ directed along z-axis is applied on the whole model.

A two-level mesh-sensitivity study was conducted in Abaqus by comparing the maximum displacement obtained with a coarse discretization containing 3240 nodes and 1670 C3D8R elements and a refined discretization containing 16,860 nodes and 8592 C3D8R elements. For computational convenience, only in this validation context, an isotropic material with the following properties is considered: Young modulus $E = 99\,\mathrm{GPa}$, Poisson ratio $\nu = 0.286$, and density $\rho = 1410\,\frac{\mathrm{kg}}{\mathrm{m}^3}$. The difference in maximum displacement between the two discretizations is approximately 7%. The refined mesh was retained as a pragmatic compromise for generating a consistently discretized proof-of-concept database. This value is not interpreted as a 7% bound on local stress or failure-index errors, which can be more sensitive to discretization. Since the same refined mesh is employed for both undamaged and damaged configurations, the comparative trends and surrogate-training labels remain internally consistent, whereas the absolute failure-index values remain subject to discretization uncertainty. Accordingly, no certification-level or high-fidelity accuracy is claimed.

2.3. Damage characteristics

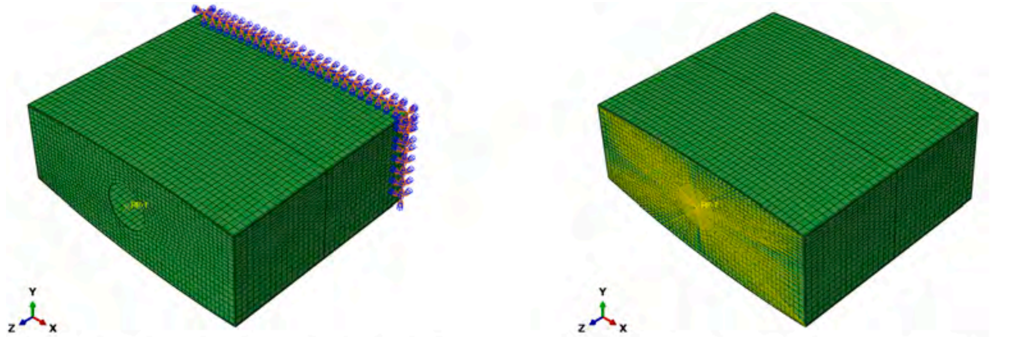
Damage is represented through prescribed local stiffness degradation within selected finite-element patches. The database records three independent descriptors: damage location, damage extent, and damage intensity. These descriptors provide a controlled parameterization for evaluating how changes in the equivalent structural stiffness affect the predicted failure indices:

- Damage location: the precise location of the damage within the wing box is documented, involving specifying coordinates and also referencing them to specific structural elements (ribs, spars, skins);
- Damage extent: the geometric details of the damage are captured, including data on the size of the damaged area (expressed in $[\mathrm{m}^2]$);
- Damage intensity: to numerically simulate the presence of damage, Young and shear elastic moduli in the selected damaged zone of the model are degraded by a percentage of their undamaged values.

In detail, damage intensity is expressed through a scalar index, defined as

$$D_I = 1 - \frac{E_{ij}^{\mathrm{dam}}}{E_{ij}^{\mathrm{und}}}, \tag{1}$$

where E_{ij}^{dam} and E_{ij}^{und} denote the degraded and pristine elastic-stiffness components, respectively, while the indices i and j identify the material directions. Four prescribed levels are considered, namely D_I = 0.20, 0.40, 0.60, 0.80. This phenomenological representation captures



(a) Fixed constraint applied on the external surface of Rib 3.

(b) Distributed static loads applied on the external surface of Rib 1.

Fig. 3. Boundary conditions and external loads applied on the reference wing box.

the averaged residual-stiffness consequence of intralaminar degradation, but it does not reproduce the initiation mechanism, crack morphology, progressive failure process, or possible delamination associated with a physical impact or fatigue event. It is therefore used to generate controlled numerical damage states rather than to simulate the complete physical damage-evolution process.

2.4. Adopted failure criteria

This work examines five classical FIs for composite materials, each from a different perspective, to assess the damage state of the component under analysis. These failure criteria are defined in detail as follows.

- The Max Stress Criterion identifies composite material failure caused by three possible modes of loading: longitudinal failure, transverse failure, and shear failure:
  - Longitudinal failure occurs whenever $\sigma_{11} \geq S_{11}^+$ or $\sigma_{11} \leq S_{11}^-$;
  - Transverse failure occurs whenever $\sigma_{22} \geq S_{22}^+$ or $\sigma_{22} \leq S_{22}^-$;
  - Longitudinal shear failure occurs whenever $|\sigma_{12}| \geq |S_{12}|$.

  FI = Max. Absolute Value of $\left(\frac{\sigma_{11}}{S_{11}^+}, \frac{\sigma_{11}}{S_{11}^-}, \frac{\sigma_{22}}{S_{22}^+}, \frac{\sigma_{22}}{S_{22}^-}, \frac{\sigma_{12}}{S_{12}}\right) \geq 1$.
- The Tsai-Wu Criterion is a quadratic, interactive stress-based criterion that identifies failure but does not distinguish between different failure modes. Failure occurs whenever the following condition is satisfied: $F_1\sigma_{11} + F_2\sigma_{22} + F_{11}\sigma_{11}^2 + F_{22}\sigma_{22}^2 + F_{66}\sigma_{12}^2 + 2F_{12}\sigma_{11}\sigma_{22} \geq 1$. The various coefficients F_i and F_{ij} of the Tsai-Wu criterion are defined in terms of known/measured strengths of the composite material: $F_1 \equiv \frac{1}{S_{11}^+} - \frac{1}{S_{11}^-}$; $F_2 \equiv \frac{1}{S_{22}^+} - \frac{1}{S_{22}^-}$; $F_{11} \equiv \frac{1}{S_{11}^+S_{11}^-}$; $F_{22} \equiv \frac{1}{S_{22}^+S_{22}^-}$; $F_{66} \equiv \frac{1}{S_{12}S_{12}}$. The interaction coefficient F_{12} is computed as $F_{12} = f\sqrt{F_{11}F_{22}}$, where f is a user-specified constant, $-0.5 \leq f \leq 0$.
- The Tsai-Hill criterion is a quadratic, interactive stress-based criterion that identifies failure but does not distinguish between different failure modes. Failure occurs whenever the following condition is satisfied: $\frac{\sigma_{11}^2}{S_{11}^2} - \frac{\sigma_{11}\sigma_{22}}{S_{11}^2} + \frac{\sigma_{22}^2}{S_{22}^2} + \frac{\sigma_{12}^2}{S_{12}^2} \geq 1$.
- The Hashin criterion identifies four different failure modes for the composite material. The four modes are tensile fiber failure, compressive fiber failure, tensile matrix failure, and compressive matrix failure.
  - If $\sigma_{11} \geq 0$, the Tensile fiber failure criterion is: $\left(\frac{\sigma_{11}}{S_{11}^+}\right)^2 + \alpha\left(\frac{\sigma_{12}^2+\sigma_{13}^2}{S_{12}^2}\right) \geq 1$;
  - If $\sigma_{11} < 0$, the compressive fiber failure criterion is: $\left(\frac{\sigma_{11}}{S_{11}^-}\right)^2 \geq 1$;
  - If $\sigma_{22} + \sigma_{33} \geq 0$, the tensile matrix failure criterion is: $\frac{(\sigma_{22}+\sigma_{33})^2}{S_{22}^{+2}} + \frac{\sigma_{23}^2-\sigma_{22}\sigma_{33}}{S_{23}^2} + \frac{\sigma_{12}^2+\sigma_{13}^2}{S_{12}^2} \geq 1$;
  - If $\sigma_{22} + \sigma_{33} < 0$, the compressive matrix failure criterion is: $\left[\left(\frac{S_{22}^-}{2S_{23}}\right)^2 - 1\right]\left(\frac{\sigma_{22}+\sigma_{33}}{S_{22}^-}\right) + \frac{(\sigma_{22}+\sigma_{33})^2}{4S_{23}^2} + \frac{\sigma_{23}^2-\sigma_{22}\sigma_{33}}{S_{23}^2} + \frac{\sigma_{12}^2+\sigma_{13}^2}{S_{12}^2} \geq 1$.

  The Hashin equations require a user-specified parameter α; the allowable range is $0 \leq \alpha \leq 1$, and the default value is $\alpha = 0$.
- The Christensen criterion identifies two failure modes: matrix failure and fiber failure:
  - The matrix failure criterion is: $\left(\frac{1}{S_{22}^+} - \frac{1}{S_{22}^-}\right)(\sigma_{22} + \sigma_{33}) + \frac{1}{S_{22}^+S_{22}^-}(\sigma_{22} + \sigma_{33})^2 + \frac{1}{S_{23}^2}(\sigma_{23}^2 - \sigma_{22}\sigma_{33}) + \frac{1}{S_{12}^2}(\sigma_{12}^2 + \sigma_{13}^2) \geq 1$;
  - The fiber failure criterion is: $\left(\frac{1}{S_{11}^+} - \frac{1}{S_{11}^-}\right)\sigma_{11} + \frac{\sigma_{11}^2}{S_{11}^+S_{11}^-} \geq \mathbf{1}$.

2.5. Finite element analyses for structural health data generation

The finite-element model is employed as a controlled virtual data generator. For each prescribed damage configuration, the stress field and five classical failure indices are evaluated throughout the wing-box model. The resulting labels describe the response of the adopted numerical representation; consequently, the present study evaluates the ability of machine-learning models to emulate that response. Experimental calibration, model updating, and simulation-to-measurement transfer are required before operational application.

The parametric campaign comprises one undamaged configuration and 112 damaged configurations obtained by varying the affected structural region, damaged-patch extent, and four stiffness-reduction levels. Compared with the previously published material-oriented benchmark [19], the present study focuses on an element-level, single-material, damage-oriented database; it also broadens the model comparison to Random Forest, XGBoost, LightGBM, and a multilayer perceptron and explicitly examines prediction errors within the upper failure-index range (see Table 3).

The results are extracted as the six direct stress components at each centroid of the numerical model elements, which are used to identify maximum FIs on a case-by-case basis, as summarized in Figs. 4 and 5. Since this parametric analysis aims to evaluate the ANN's ability to elucidate the dependence of the five selected FIs on intralaminar damage conditions, the database is structured accordingly. The results are also presented on a case-by-case basis, with respect to the element whose FIs vary the most relative to the undamaged configuration. Such findings are summarized in Figs. 6 and 7.

It is interesting to note that, for both investigation approaches, the most varied and interesting cases appear to be those from 81 to 96,

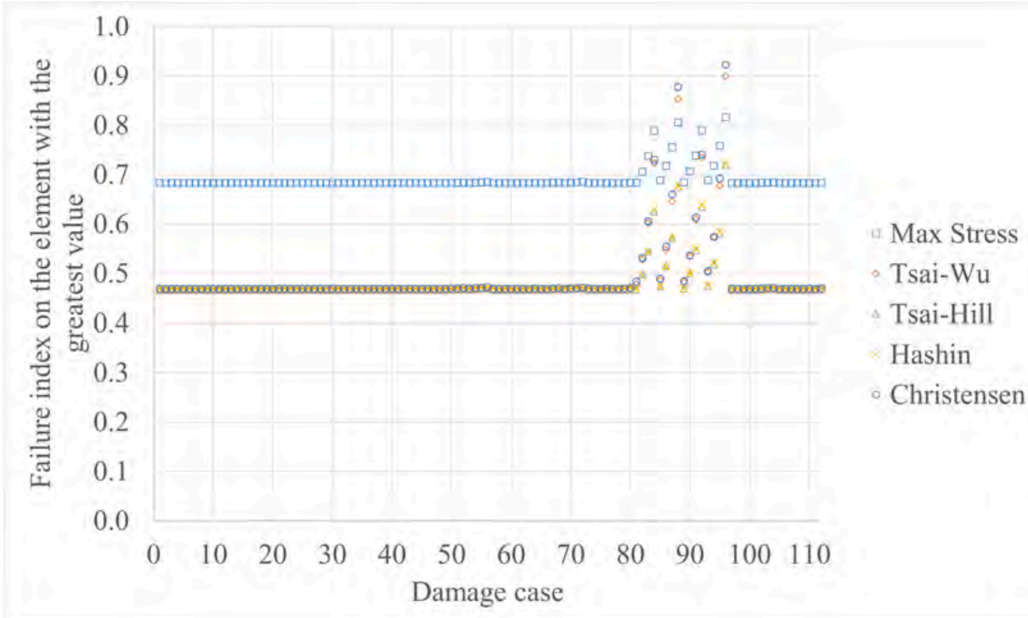


Fig. 4. Five FIs on the element with the greatest value, computed case by case based on the whole model failure map.

Table 3
Damage cases for the parametric analysis aiming at a damage-oriented database.

Case	Position	Centroid coordinates [m]			Extent [m^2]
		x	y	z	
1	Rib	4.21E-01	−3.60E-01	0.00E+00	4.24E-04
5	Rib	4.21E-01	−3.60E-01	0.00E+00	1.70E-03
9	Rib	7.45E-01	−3.42E-01	0.00E+00	4.24E-04
13	Rib	7.45E-01	−3.42E-01	0.00E+00	1.70E-03
17	Rib	4.21E-01	−3.60E-01	2.63E-01	4.24E-04
21	Rib	4.21E-01	−3.60E-01	2.63E-01	1.70E-03
25	Rib	7.45E-01	−3.42E-01	2.63E-01	4.24E-04
29	Rib	7.45E-01	−3.42E-01	2.63E-01	1.70E-03
33	Rib	4.21E-01	−3.60E-01	5.25E-01	4.24E-04
37	Rib	4.21E-01	−3.60E-01	5.25E-01	1.70E-03
41	Rib	7.45E-01	−3.42E-01	5.25E-01	4.24E-04
45	Rib	7.45E-01	−3.42E-01	5.25E-01	1.70E-03
49	Skin	5.95E-01	−4.72E-01	1.33E-01	3.20E-03
53	Skin	5.95E-01	−4.72E-01	1.33E-01	7.20E-03
57	Skin	5.95E-01	−4.72E-01	3.95E-01	3.20E-03
61	Skin	5.95E-01	−4.72E-01	3.95E-01	7.20E-03
65	Skin	5.94E-01	−2.37E-01	1.33E-01	3.20E-03
69	Skin	5.94E-01	−2.37E-01	1.33E-01	7.20E-03
73	Skin	5.81E-01	−2.37E-01	3.95E-01	3.20E-03
77	Skin	5.81E-01	−2.37E-01	3.95E-01	7.20E-03
81	Spar	2.97E-01	−3.67E-01	1.33E-01	6.28E-04
85	Spar	2.97E-01	−3.67E-01	1.33E-01	2.51E-03
89	Spar	2.97E-01	−3.67E-01	3.95E-01	6.28E-04
93	Spar	2.97E-01	−3.67E-01	3.95E-01	2.51E-03
97	Spar	8.79E-01	−3.34E-01	1.33E-01	6.28E-04
101	Spar	8.79E-01	−3.34E-01	1.33E-01	2.51E-03
105	Spar	8.79E-01	−3.34E-01	3.95E-01	6.28E-04
109	Spar	8.79E-01	−3.34E-01	3.95E-01	2.51E-03

i.e., those involving damage to a specific spar, thereby identifying that component as the most critical from a structural failure perspective.

2.6. Machine-learning surrogates

The element-level database was used to train and compare several regression surrogates for each failure index. Random Forest, XGBoost, and LightGBM were selected as candidate models because tree ensembles are well-suited to heterogeneous tabular data and non-linear interactions. A multilayer perceptron was retained as a benchmark to assess whether a generic neural architecture and a broader feature set confer an advantage on the available dataset. The ANN is therefore treated as a comparative model rather than as the preferred solution.
The workflow comprises exploratory data analysis, feature engineering, model design and training, and validation. Hyperparameter selection and validation are connected iteratively.

2.6.1. Exploratory data analysis

The Exploratory Data Analysis phase is a crucial step to uncover trends and insights embedded in the dataset. Descriptive statistics and visual inspection were used to assess the distributions of the input and output variables.

Table 4 summarizes the mean and standard deviation of the output variables grouped by damage type, while Table 5 reports the same statistics grouped by categorical position. Scatter plots in Fig. 8 show the relationship between structural damage variables and the stress tensor, while Fig. 9 replaces the stress tensor with the FI. A color code encodes the categorical position.

Since no clear linear dependencies emerged, the prediction task was reformulated to use the FIs directly as target variables rather than predicting the stress tensor as an intermediate step.
The linear association between two random variables is commonly assessed through the Pearson correlation coefficient, r_{xy}. This statistic provides a normalized measure of both the intensity and the direction of their linear dependence, and it is formally expressed as:

$$r_{xy} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X, \sigma_Y} \tag{2}$$

Here, $\mathrm{Cov}(X,Y)$ represents the covariance between X and Y, while σ_X and σ_Y denote their respective standard deviations. The coefficient ranges between -1 and 1: values close to 1 indicate a strong positive linear dependence, values close to -1 correspond to a strong negative linear dependence, and values near 0 suggest the absence of a linear relationship. It should be noted, however, that r_{xy} only characterizes linear dependencies and does not capture potential non-linear interactions between variables, which may require alternative measures of association.

The analysis was extended to the FIs by examining the correlation matrices for each damage case, considering only the measurement element corresponding to the maximum FI value. This approach was

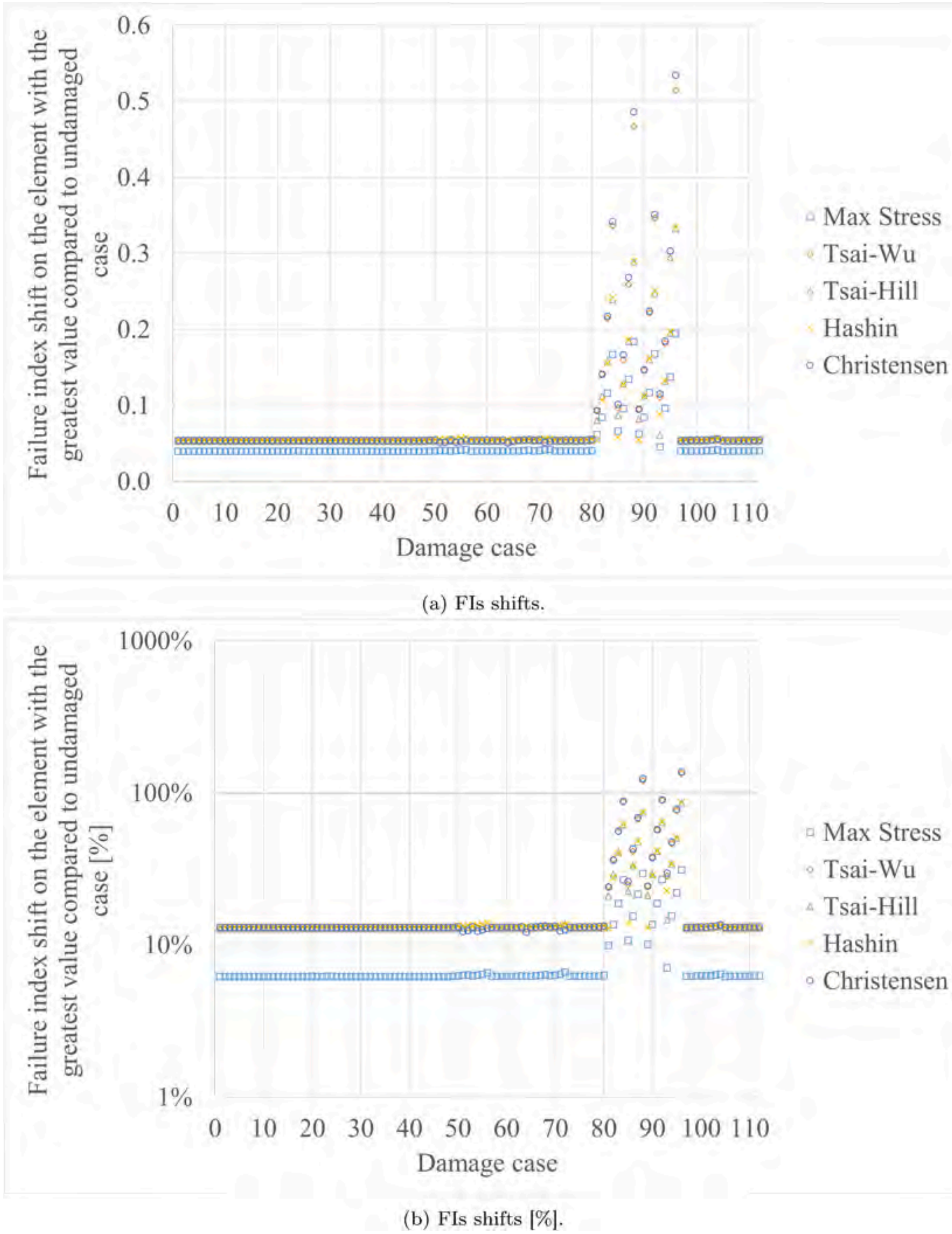


Fig. 5. Five FIs shifts, computed case by case on the element with the greatest value, identified based on the whole model failure map.

adopted to highlight the most critical structural points. Each resulting matrix, therefore, represents 112 observed nodes. The correlation analysis revealed a consistent pattern: for all FIs, a negative correlation was identified with the x-component of the stressed node, while positive correlations were observed with both the spatial position and the applied load intensity, as can be seen for the Max Stress index, shown in Fig. 10.

A detailed analysis of maximum FI distributions, shown in Fig. 11, indicated that only damage cases with IDs between 81 and 96 exhibited significantly higher values, consistently associated with the Spar position and nodes 1402 and 1522. To further investigate this dependency, a sub-dataset of 32 records was extracted by filtering on Spar position. The centroid of each measurement group was computed and incorporated to assess its relation with stressed nodes. Results demonstrated that, for cases 81–96, the x-coordinate of the centroid was the primary driver of variation in FIs, as can be seen in Fig. 12.

2.6.2. *Feature engineering*

Feature engineering plays a central role in defining the problem space and ensuring effective predictions of structural health monitoring. This phase involves identifying relevant variables, applying scaling techniques, and selecting variables based on domain knowledge.

Categorical Feature Encoding. The categorical feature *Position* was encoded using one-hot encoding to prevent unintended ordinal relationships. Although this can increase dimensionality, it proved effective at ensuring robust compatibility with decision-tree ensembles and neural networks.

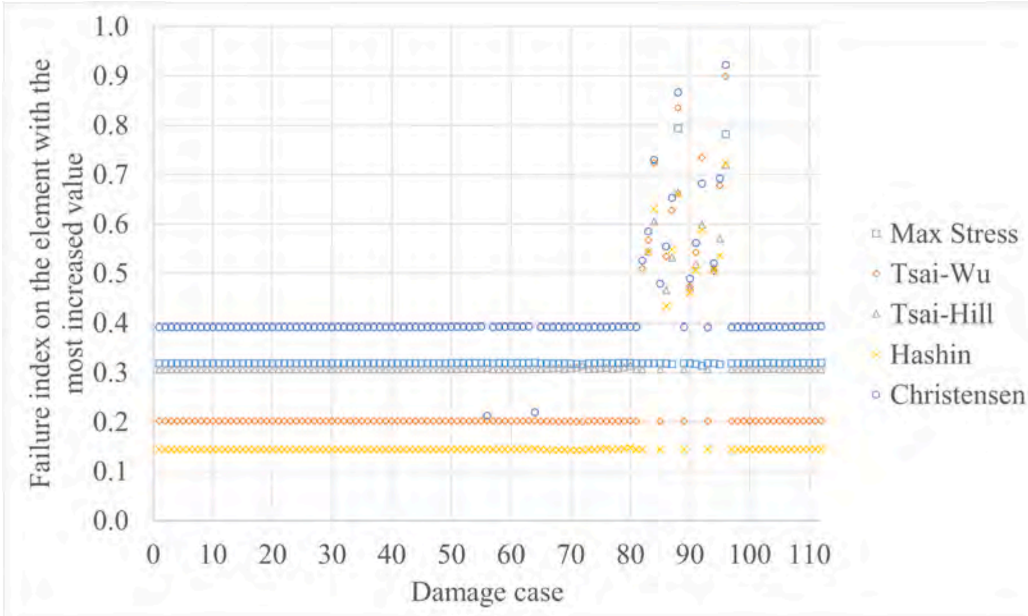


Fig. 6. Five FIs on the element with the most increased value, computed case by case based on the whole model failure map.

Feature Scaling. Standard Gaussian-based normalization was discarded due to multi-modal data distributions. Instead, min-max scaling was applied before feeding data into the ANN, as it accelerates gradient-based optimization and prevents dominance of high-magnitude variables. To avoid bias, the scaling parameters were fitted exclusively to the training data before the data were split into validation and test subsets. For ensemble methods, scaling was unnecessary because they are inherently scale-invariant.

Feature Selection. Both domain expertise and model suitability guided feature selection. ML models were primarily trained on damage-related variables and centroid information to reduce computational overhead. Conversely, the ANN was trained on a broader feature set, including material characteristics and nodal coordinates of the 8-node measurement element. This choice was motivated by observed correlations between nodal coordinates and FIs.

2.6.3. Model design and training

In this study, different machine learning strategies were explored to model the relationship between structural damage descriptors and FIs. Two main classes of models were developed: tree-based ensemble methods (Random Forest, Extreme Gradient Boosting, and Light Gradient Boosting) and an ANN. Ensemble methods were selected for their robustness to low-variance data, while the neural network was motivated by the dataset's high intrinsic dimensionality and non-linear dependencies. The ANN is retained as a controlled benchmark for comparison with common neural approaches in SHM, whereas the subsequent results identify the ensemble models as the more suitable predictors for the present dataset. The ensemble models were implemented using `scikit-learn`, whereas the ANN was developed in `TensorFlow/Keras`. A separate model was trained for each FI.

Ensemble Methods. Random Forest is an ensemble algorithm that constructs a set of decision trees and aggregates their predictions to improve accuracy and reduce overfitting. Diversity among trees is introduced through bootstrapped samples and random feature selection. The Random Forest Regressor was optimized by tuning:

- `n_estimators`: [50, 100, 200]; number of trees in the forest
- `max_depth`: [None, 10, 20, 30]; maximum depth of each tree
- `min_samples_split`: [2, 5, 10]; minimum samples required to split an internal node
- `min_samples_leaf`: [1, 2, 4]; minimum samples required at a leaf node

An exhaustive grid search was employed through `GridSearchCV`, systematically exploring all parameter combinations. Each configuration was evaluated using k-fold cross-validation, and the best-performing set was retained. The primary scoring metric was `neg_mean_absolute_error`; the Pearson correlation coefficient r_{xy} was also computed.

XGBoost builds trees sequentially, with each new tree correcting residuals of the current ensemble; regularization in the objective limits overfitting. The XGB Regressor was tuned over:

- `n_estimators`: [100, 200]; number of boosting rounds
- `learning_rate`: [0.01, 0.1, 0.3]; shrinkage per boosting step
- `max_depth`: [6, 10]; maximum tree depth
- `min_child_weight`: [1]; minimum sum of Hessian in a child node
- `subsample`: [0.8, 1.0]; fraction of training instances per iteration

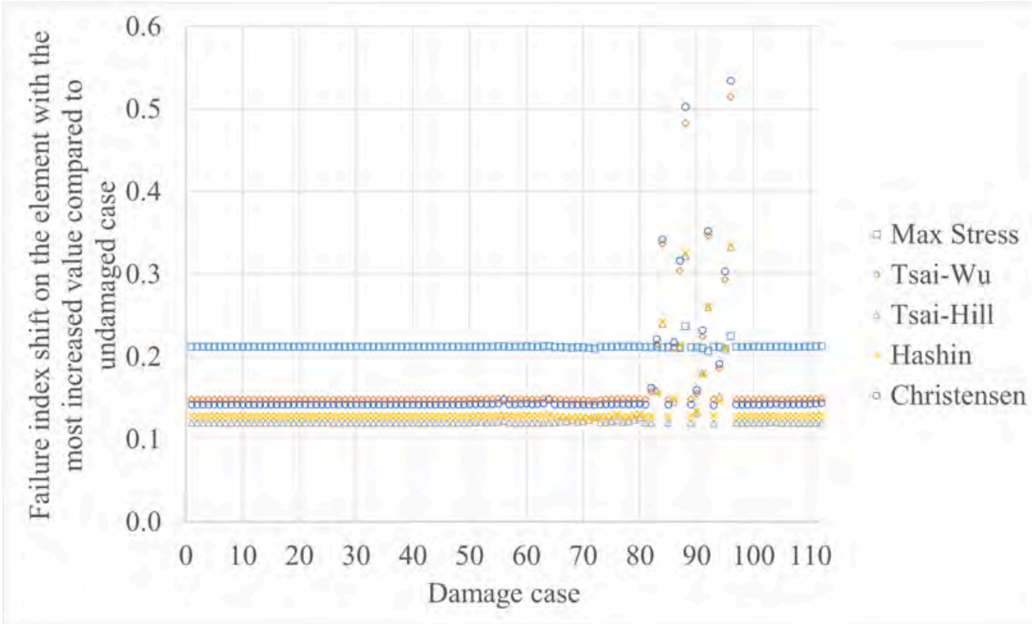
The optimal configuration was identified via grid search with cross-validation.
LightGBM employs histogram-based binning and leaf-wise growth, often enabling faster training on large datasets. The LGBM Regressor was tuned over:

- `num_leaves`: [8, 31, 63]; maximum number of leaves per tree
- `max_depth`: [-1, 3, 6]; maximum tree depth (−1 means no limit)
- `learning_rate`: [0.01, 0.05, 0.1]; shrinkage rate
- `n_estimators`: [100, 300]; number of boosting iterations
- `min_child_samples`: [10, 20, 30]; minimum samples per leaf
- `subsample`: [0.7, 0.8, 1.0]; fraction of data per iteration
- `colsample_bytree`: [0.7, 0.8, 1.0]; fraction of features per tree

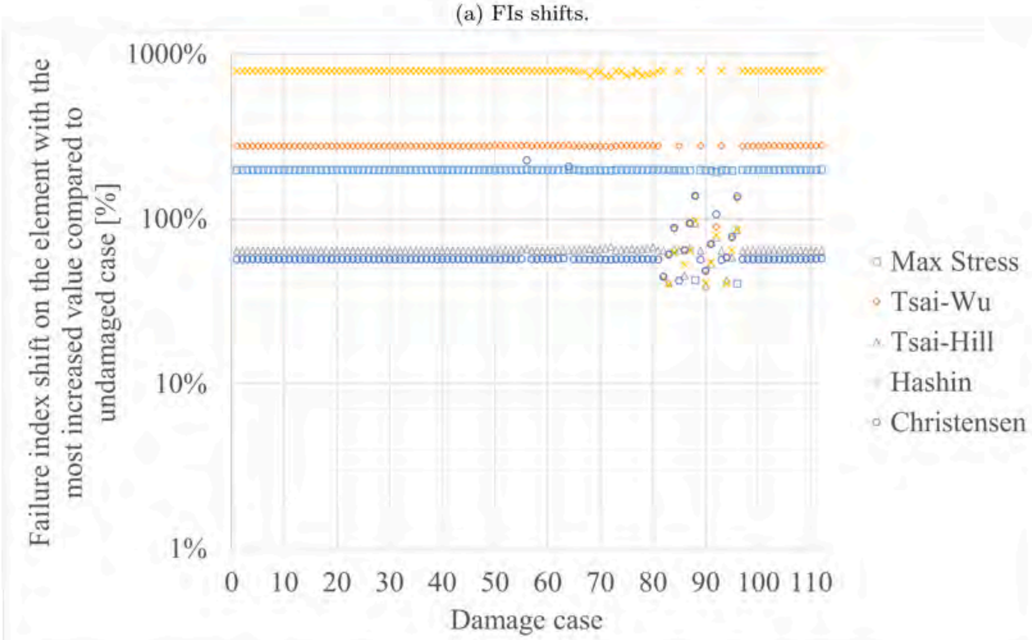
Hyperparameters were selected via `GridSearchCV`.

Artificial Neural Network. Multilayer Perceptrons (MLPs) were used to capture non-linear relations between features and FIs. The network was implemented in `TensorFlow/Keras` with the following setup:

- Data split: 70% training (673,611 samples), 20% validation (192,462), 10% test (96,231)



(a) FIs shifts.



(b) FIs shifts [%].

Fig. 7. Five FIs shifts, computed case by case on the element whose value exhibits the greatest increase compared to the undamaged configuration, identified based on the whole model failure map.

- Inputs: 51 features (solicited node coordinates: 3; measurement nodes: 24; material properties: 16; position: 3; intensity; extent; centroid of measurement group: 3)
- Epochs: 100
- Batch size: 32
- Optimizer: Adam, learning rate = 0.0005
- Loss: `mse`
- Metric: `mae`
- Architecture: 64, 32, 16, 1 (three hidden layers plus output); activations: ReLU (hidden), sigmoid (output)
- Regularization: batch normalization; learning-rate scheduling; early stopping

In this architecture, the network progressively reduces dimensionality by decreasing the number of neurons at each layer (64 → 32 → 16 → 1). This pyramidal structure enables the model to extract high-level abstractions while gradually mitigating the risk of overfitting. The use of ReLU activations in the hidden layers provides non-linearity. It prevents the vanishing-gradient problem, while the sigmoid activation in the output layer is well-suited to bounded regression outputs. Regularization mechanisms such as batch normalization, adaptive learning rate scheduling, and early stopping are employed to improve convergence stability and generalization on the validation and test sets.

From a scientific perspective, this network architecture strikes a balance between model complexity and data dimensionality, ensuring sufficient representational capacity without incurring excessive overfit-

Table 4
Detailed results by damaged node (mean ± std).

(a) Normal stresses.

Damaged Node	σ_{11} [Pa]	σ_{22} [Pa]	σ_{33} [Pa]
1402	6.04e+15 ± 1.61e+16	−4.70e+16 ± 8.66e+15	−8.67e+14 ± 1.70e+15
1522	6.03e+15 ± 1.61e+16	1.36e+16 ± 8.69e+15	−8.81e+15 ± 1.70e+16
3695	5.94e+14 ± 1.61e+16	7.65e+15 ± 6.88e+15	−8.77e+15 ± 1.70e+16
5319	5.94e+15 ± 1.61e+16	7.76e+14 ± 6.88e+15	−8.77e+15 ± 1.70e+16
5351	5.94e+15 ± 1.61e+16	7.64e+15 ± 6.88e+15	−8.77e+15 ± 1.70e+16
7777	5.84e+15 ± 1.62e+16	1.37e+16 ± 6.91e+15	−8.79e+15 ± 1.72e+16
7989	5.87e+15 ± 1.62e+16	1.74e+16 ± 6.91e+15	−8.78e+15 ± 1.71e+16
8623	6.08e+14 ± 1.62e+15	−1.73e+16 ± 6.91e+15	−8.83e+15 ± 1.72e+16
8836	6.08e+15 ± 1.62e+16	1.92e+16 ± 6.91e+14	−8.84e+15 ± 1.71e+15
11303	5.94e+15 ± 1.61e+16	7.68e+15 ± 6.88e+15	−8.77e+15 ± 1.70e+15
11333	5.94e+15 ± 1.61e+16	7.53e+15 ± 6.88e+14	−8.76e+15 ± 1.70e+16
11363	5.94e+15 ± 1.61e+16	7.67e+15 ± 6.88e+15	−8.77e+15 ± 1.70e+16
15202	5.84e+15 ± 1.61e+16	1.53e+16 ± 6.99e+15	−8.81e+14 ± 1.70e+15
15282	5.85e+15 ± 1.61e+15	6.54e+15 ± 7.01e+15	−8.79e+15 ± 1.70e+16

(b) Shear stresses.

Damaged Node	σ_{12} [Pa]	σ_{13} [Pa]	σ_{23} [Pa]
1402	4.81e+15 ± 1.19e+16	−5.28e+16 ± 6.60e+16	3.60e+15 ± 6.28e+14
1522	4.84e+15 ± 1.19e+16	−5.29e+16 ± 6.60e+15	3.49e+15 ± 6.27e+15
3695	4.82e+14 ± 1.19e+16	−5.29e+15 ± 6.53e+15	3.55e+15 ± 6.26e+15
5319	4.82e+15 ± 1.19e+16	−5.29e+15 ± 6.53e+14	3.55e+15 ± 6.26e+15
5351	4.82e+15 ± 1.19e+16	−5.29e+15 ± 6.53e+15	3.55e+15 ± 6.26e+15
7777	4.83e+15 ± 1.19e+16	−5.24e+15 ± 6.54e+15	3.53e+15 ± 6.27e+15
7989	4.82e+15 ± 1.19e+16	−5.28e+15 ± 6.56e+16	3.59e+15 ± 6.27e+15
8623	4.83e+15 ± 1.19e+16	−5.43e+16 ± 6.54e+15	3.51e+15 ± 6.26e+15
8836	4.82e+15 ± 1.19e+16	−5.23e+15 ± 6.55e+15	3.58e+15 ± 6.27e+14
11303	4.82e+15 ± 1.19e+16	−5.29e+14 ± 6.53e+15	3.55e+15 ± 6.26e+14
11333	4.82e+15 ± 1.19e+14	−5.29e+15 ± 6.53e+15	3.55e+15 ± 6.26e+15
11363	4.82e+15 ± 1.19e+16	−5.29e+15 ± 6.53e+15	3.55e+15 ± 6.26e+14
15202	4.83e+16 ± 1.19e+16	−5.30e+15 ± 6.52e+16	3.54e+16 ± 6.26e+15
15282	4.82e+15 ± 1.19e+16	−5.30e+15 ± 6.52e+15	3.56e+16 ± 6.26e+15

(c) Failure criteria.

Damaged Node	Max Stress [-]	Tsai-Wu [-]	Tsai-Hill [-]	Hashin [-]	Christensen [-]
1402	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
1522	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
3695	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
5319	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
5351	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
7777	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
7989	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
8623	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
8836	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
11303	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
11333	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
11363	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
15202	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
15282	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10

ting. The sequential reduction of neurons mirrors a feature-compression mechanism, in which redundant or weakly informative signals are filtered out as the network propagates toward the output layer.

Additionally, two alternative architectural variants were explored. In the first, an extra initial layer of 128 neurons was added before the original input layer. However, this modification led to a substantial deterioration in performance, as the model's increased complexity was not justified by the dataset's relatively low variance, resulting in overfitting. In the second variant, the first hidden layer of 64 neurons was removed, reducing the architecture to [32, 16, 1]. This simplification also adversely affected performance, as the network's reduced representational capacity was insufficient to adequately capture the data's complex relationships. These results indicate that the proposed baseline architecture strikes an appropriate balance between complexity and generalization.

2.6.4. Validation and evaluation

For the training stage, two distinct data partitioning strategies were adopted. First, a conventional 70%-30% split was used, with 70% of the data allocated to training and 30% to testing. For artificial neural networks, the test portion was further divided into validation (20%) and test (10%) subsets to enable model monitoring during training.

To guarantee a balanced representation of critical conditions, the variables associated with FI values above the previously defined threshold were distributed proportionally across all subsets, following the percentages reported in Fig. 13. The remaining FI values were clustered, and representative samples were drawn to complete the training, validation, and test partitions. This stratified procedure ensured that both standard and extreme structural responses were consistently represented in the datasets. The threshold-based stratification was specifically introduced to prevent the high-failure-index regime from being excluded from the hold-out evaluation, since this regime is the most relevant one for conservative structural health screening.

The overall training methodology was designed to be systematic and rigorous, with preprocessing strategies and hyperparameter optimization tailored to each regression model. Hyperparameters were tuned us-

Table 5
Summary by position (mean ± std).

(a) Damage extent and intensity.

Position	Extent [m^2]	Intensity [-]
Rib	1.06e-03 ± 6.36e-04	0.50 ± 0.22
Skin	5.20e-03 ± 2.00e-03	0.50 ± 0.22
Spar	1.57e-03 ± 9.42e-04	0.50 ± 0.22

(b) Normal stresses.

Position	σ_{11} [Pa]	σ_{22} [Pa]	σ_{33} [Pa]
Rib	5.94e+15 ± 1.61e+16	7.65e+14 ± 6.88e+15	−8.77e+15 ± 1.70e+16
Skin	5.97e+15 ± 1.62e+16	7.85e+15 ± 6.91e+15	−8.81e+14 ± 1.72e+16
Spar	5.94e+15 ± 1.61e+15	7.67e+15 ± 7.88e+15	−8.77e+15 ± 1.69e+16

(c) Shear stresses.

Position	σ_{12} [Pa]	σ_{13} [Pa]	σ_{23} [Pa]
Rib	4.82e+15 ± 1.19e+16	−5.29e+15 ± 6.53e+15	3.55e+15 ± 6.26e+15
Skin	4.83e+15 ± 1.19e+16	−5.30e+16 ± 6.55e+15	3.55e+15 ± 6.27e+15
Spar	4.83e+15 ± 1.19e+16	−5.29e+15 ± 6.53e+15	3.55e+15 ± 6.26e+15

(d) Failure criteria.

Position	Max Stress [-]	Tsai-Wu [-]	Tsai-Hill [-]	Hashin [-]	Christensen [-]
Rib	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
Skin	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10
Spar	0.10 ± 0.17	0.05 ± 0.10	0.05 ± 0.10	0.05 ± 0.10	0.06 ± 0.10

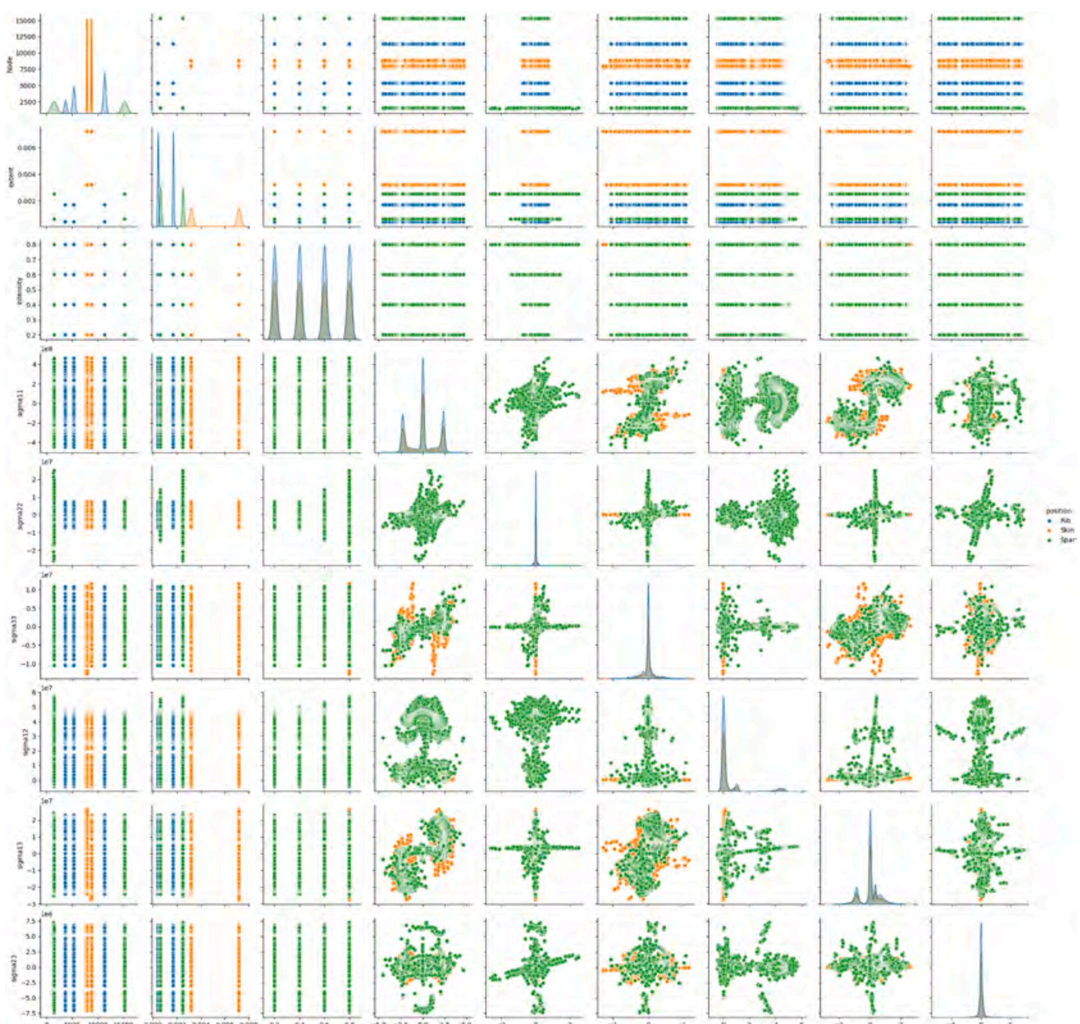


Fig. 8. Scatter plots of damage variables versus stress tensor, color-coded by categorical position.

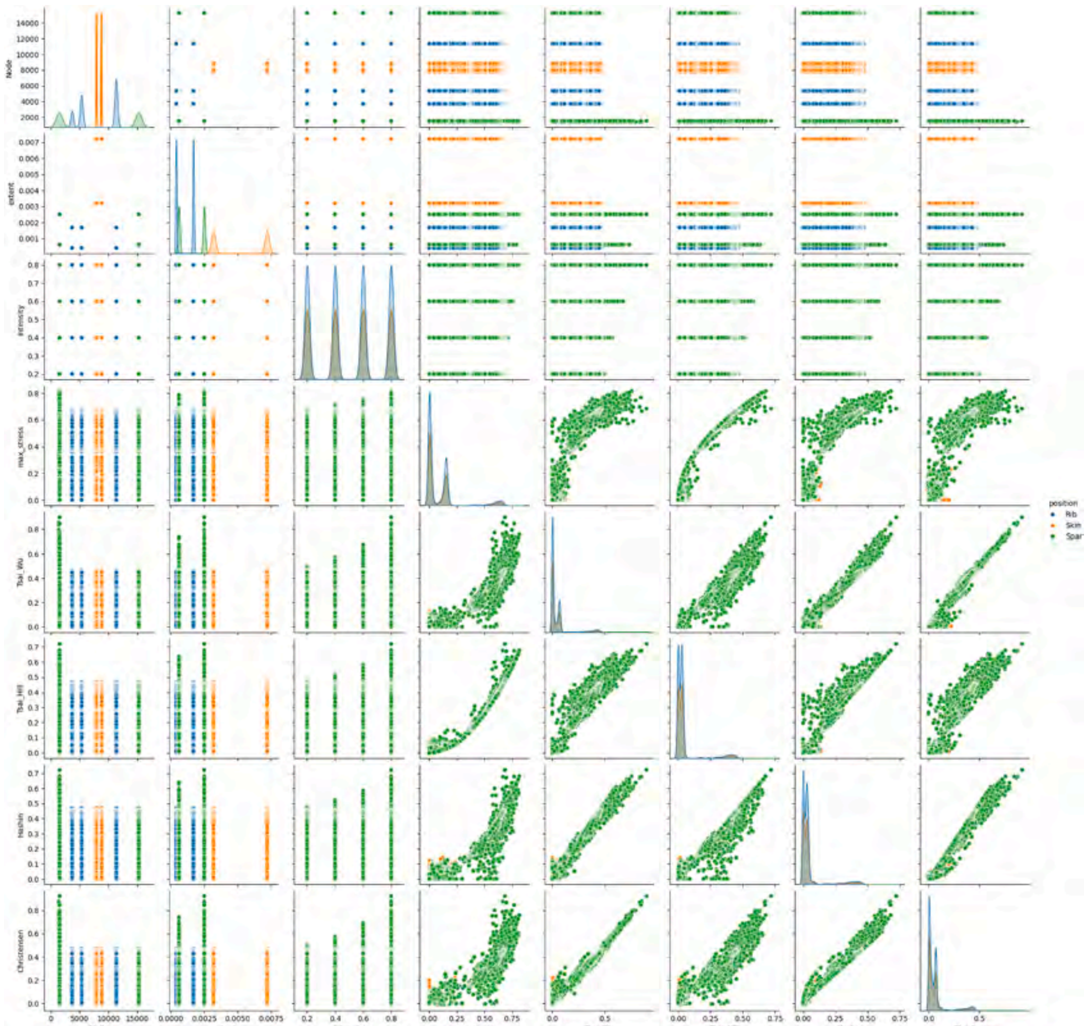


Fig. 9. Scatter plots of damage variables versus FIs, color-coded by categorical position.

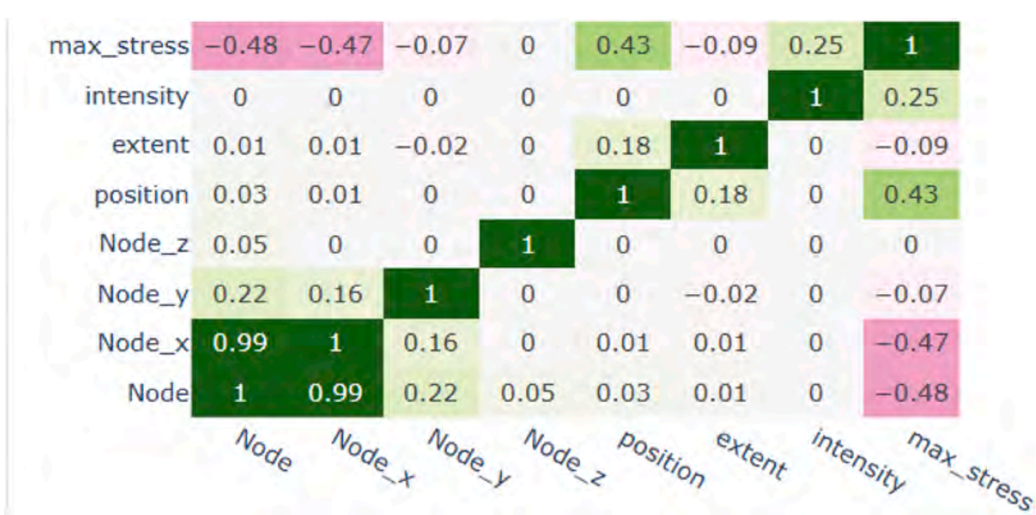


Fig. 10. Correlation matrix between structural features and Max Stress.

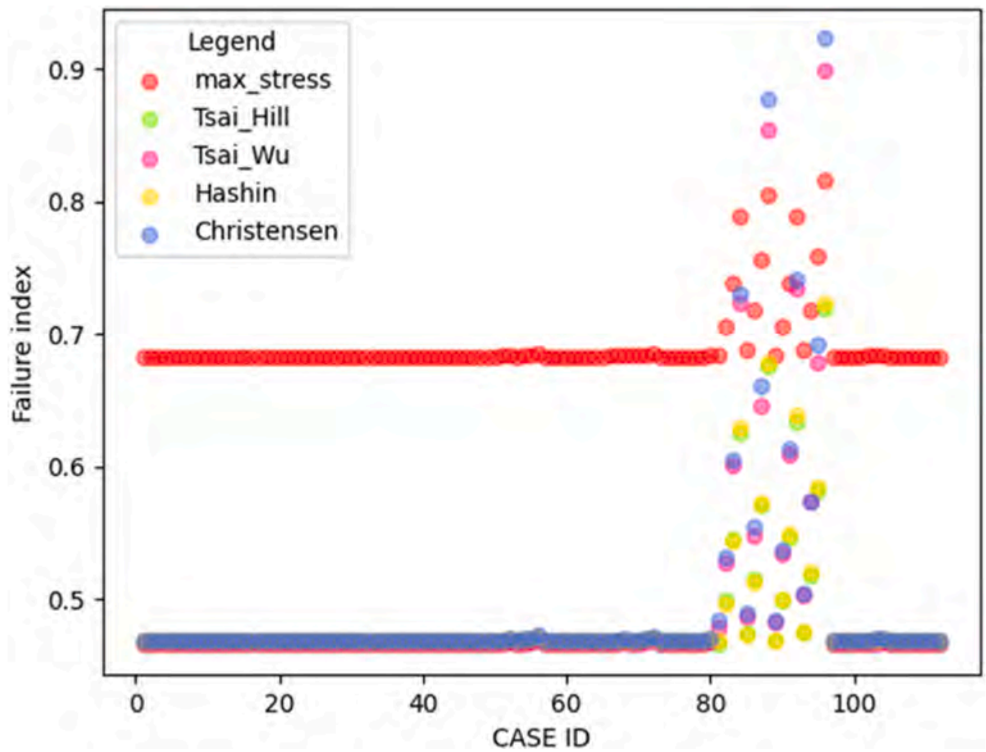


Fig. 11. Maximum FI values distribution over damage cases. Cases 81–96 are highlighted by the concentration of high-index responses associated with the spar region.

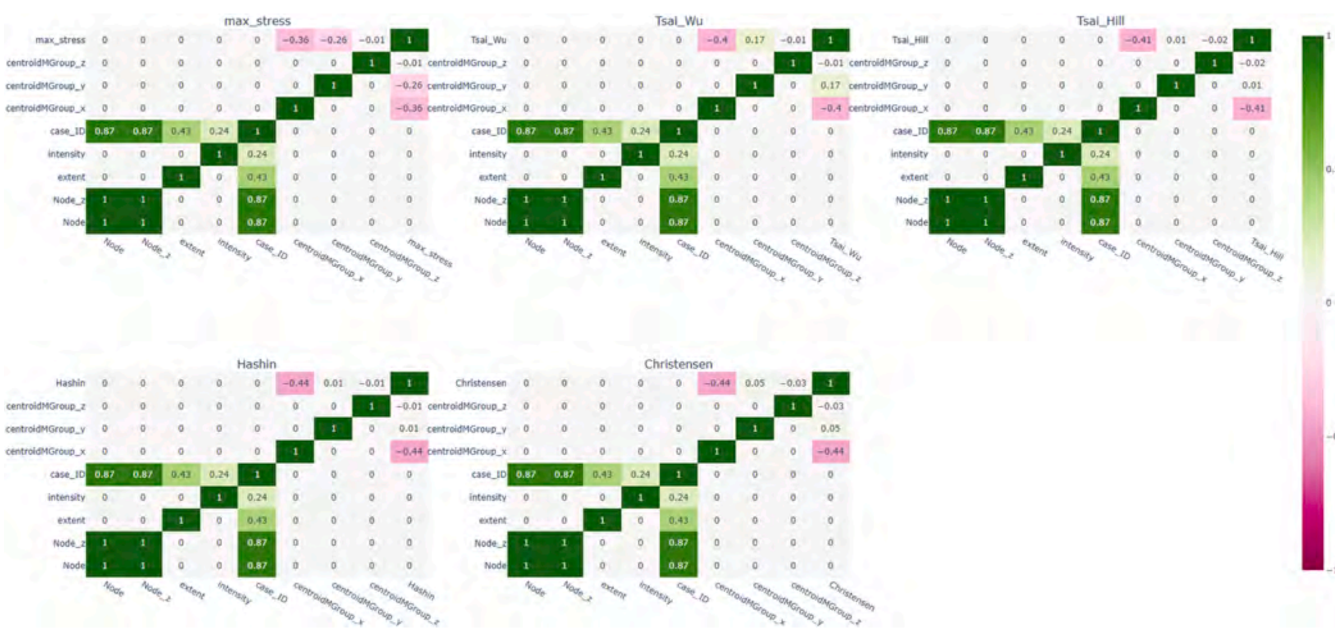


Fig. 12. Correlation matrices of the spar highest failure values.

ing an exhaustive grid search, exploring all combinations within predefined ranges.

In addition to the hold-out split, a 5-fold cross-validation procedure was employed to assess model generalization and reduce selection bias. At each iteration, the dataset was partitioned into five subsets: the model was trained on four folds and validated on the remaining one. This rotating process provided a more reliable estimate of the model's predictive capacity on unseen data, independent of the specific training subset.

The Root Mean Squared Error (RMSE) was selected as the primary evaluation metric, owing to its sensitivity to large deviations between predictions and observations. RMSE is mathematically expressed as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{3}$$

where y_i denotes the observed values, $\hat{y}_i$ the predicted values, and n the number of samples. The use of RMSE ensured that significant errors were penalized more heavily, a property particularly valuable in structural health monitoring, where outlier deviations may indicate critical damage. Furthermore, because RMSE is expressed in the same units as

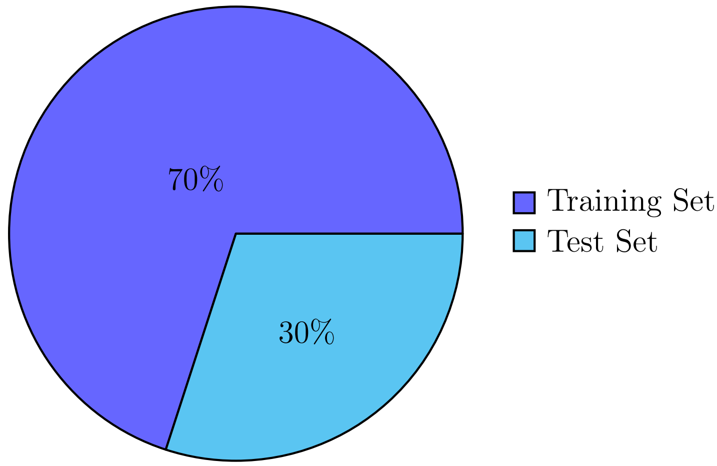


Fig. 13. Pie chart of training set and test set.

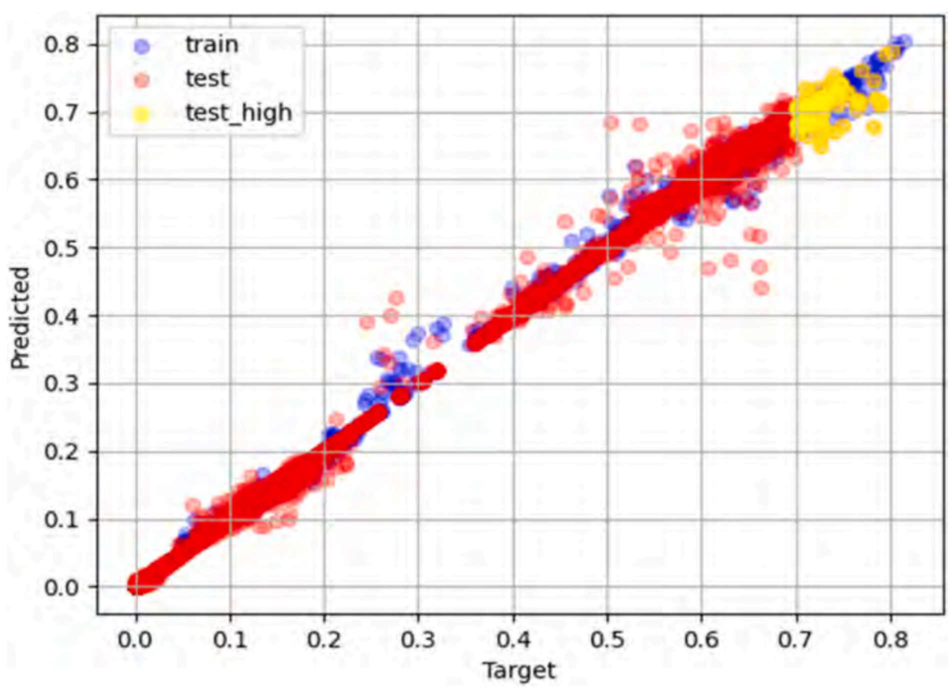


Fig. 14. Scatter plot of Max Stress - random forest prediction versus target.

the target variable, it allowed straightforward interpretation and direct comparison across different algorithms. In addition to RMSE, supplementary metrics—including Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Bias Error (MBE)—were also computed for the train-test split evaluation, providing a more complete picture of model performance. However, RMSE alone was retained as the definitive criterion, serving as the reference for comparative assessments across models.

3. Results and discussion

The results are presented for both algorithmic families, with a detailed analysis of each. Root Mean Squared Error (RMSE) is adopted as the primary performance metric because it preserves the scale of the target variable. Since the FIs lie in $[0, 1]$, RMSE values in the 0.10–0.15 range are considered good, while values below 0.30 remain acceptable for this study. Where relevant, complementary measures - Mean Squared Error (MSE), Mean Absolute Error (MAE), and Mean Bias Error (MBE) - are also reported to provide a more complete assessment of model performance.

3.1. Ensemble methods

Fig. 14 illustrates the comparison between predictions and target values for the Random Forest model trained on the `max_stress` FI. Among all the developed models, this configuration yielded the most accurate predictions. Samples are color-coded: blue for training data, red for test data, and yellow for test samples exceeding the critical threshold, i.e., FI > 0.7.

Ideally, all points should lie on the 45° parity line, which would indicate a perfect match between predicted and actual values. Deviations from the diagonal represent prediction errors, with larger distances reflecting greater inaccuracies. For low FI values—more densely expressed in the training set—the model provides highly accurate predictions. However, as the FI increases, the dispersion around the parity line also increases, indicating greater uncertainty in critical regions.

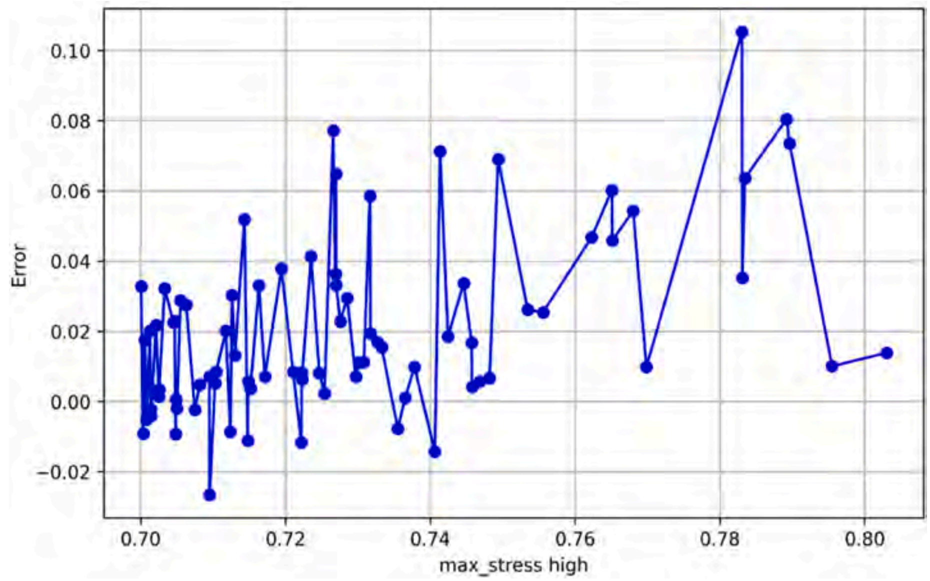


Fig. 15. Max Stress - random forest prediction error on high value.

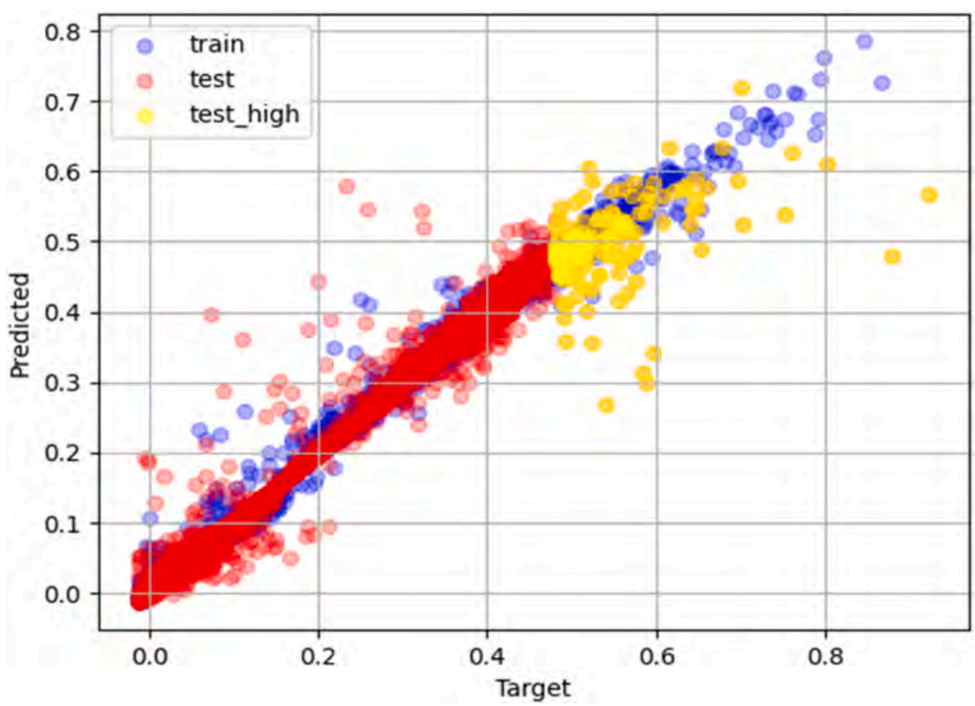


Fig. 16. Scatter plot of Christensen - random forest prediction versus target.

Cross-validation errors, reported for each fold, exhibit minimal variability, demonstrating the model's ability to generalize across unseen data. Furthermore, the alignment of the training and test performance curves confirms that the Random Forest model does not overfit.

A detailed error analysis, with error calculated as $true - predicted$, restricted to samples above the critical threshold (`max_stress` > 0.7), is shown in Fig. 15. The results indicate that prediction errors increase with the magnitude of FI. Additionally, errors are inversely proportional to sample density: in the interval [0.70-0.75], where data are more numerous, errors are higher, whereas for FI > 0.75 fewer data points are available, leading to larger variability. Since the error was computed as *actual value minus predicted value*, negative values indicate overestimation, whereas positive errors indicate underestimation. The prevalence of positive errors suggests that the model tends to underestimate FI, a bias attributable to the dataset's higher representativeness of low-damage cases. Improving the dataset with more high-damage instances could mitigate this imbalance. For this reason, the high-index subset is not interpreted through global accuracy alone. Table 7 reports dedicated error measures for samples above the threshold and shows that the Max Stress predictor remains the most reliable in the critical range, whereas the Christensen index remains the most conservative limitation of the current feature set. Future enrichment should therefore focus on spar-dominated and high-index configurations rather than uniformly adding low-damage cases.

Tables 6 and 7 summarize the metrics obtained for all FIs modeled using a Random Forest. Comparable patterns were observed across other indices, with particularly weak performance on the Christensen crite-

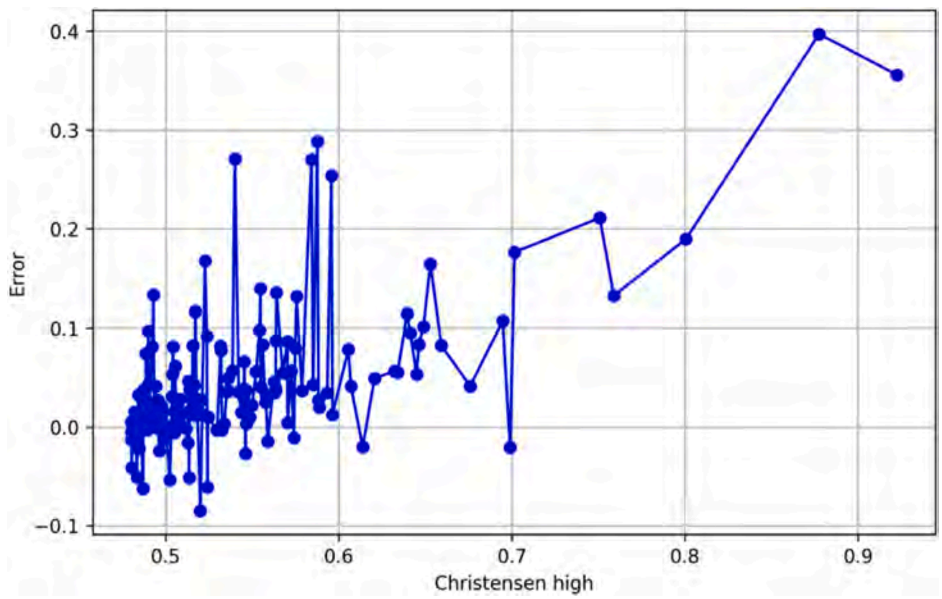


Fig. 17. Christensen - random forest prediction error on high value.

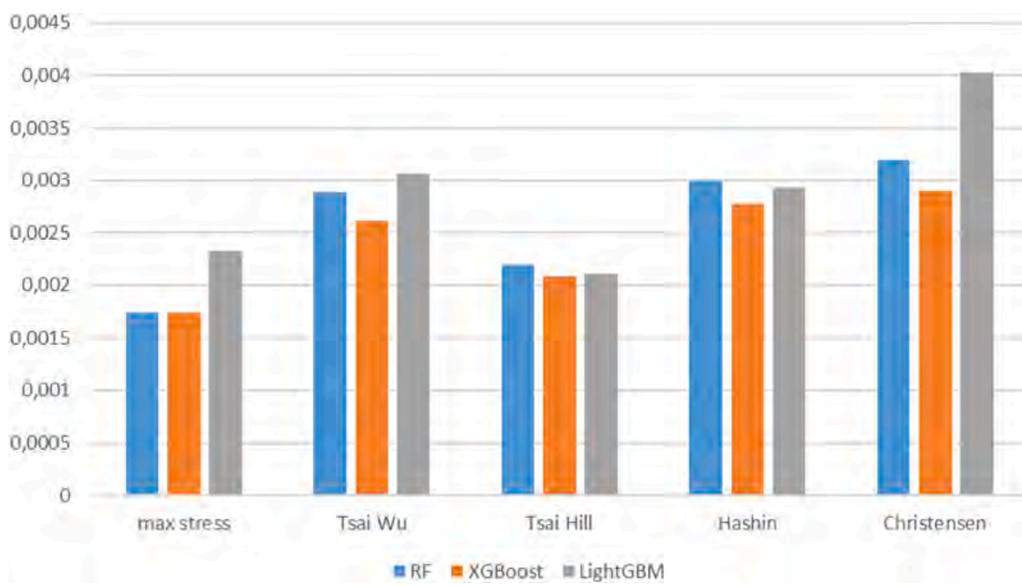


Fig. 18. Comparison of learning methods RMSE performances on test set.

rion. Figs. 16 and 17 present analogous scatter plots for Christensen, in which prediction accuracy is significantly reduced, resulting in the worst performance.

Figs. 18 and 19 report a comparative overview of the mean cross-validation error of each ensemble model (Random Forest, XGBoost, and LightGBM), respectively, on all and high index values. Among them, XGBoost consistently achieves the lowest errors, even though its predictions degrade near high FI values, which are of particular interest for structural failure mechanisms. The relative ranking of models remains stable across indices: `max_stress` consistently ranks as the best predictor, whereas Christensen ranks last. Notably, the model trained on `max_stress` maintains acceptable performance even for critical cases, supporting its suitability as a foundation for constructing a digital twin of the structure [21].

3.2. *Artificial neural networks*

Fig. 20 depicts the loss curves obtained during ANN training for both the training and validation sets. The validation set, composed of data unseen during training, was used to assess the model's generalization capability. Although the training was configured for 100 epochs, the training was terminated early by the early-stopping regularizer. Early stopping monitors the validation loss and interrupts training once the monitored metric stops improving for a specified patience period (5 epochs in this case). This mechanism prevents overfitting and reduces computational costs by halting training before convergence to a suboptimal solution. The loss trajectories do not reveal the typical divergence between training and validation losses that would indicate overfitting, suggesting that the ANN was not memorizing the training data.

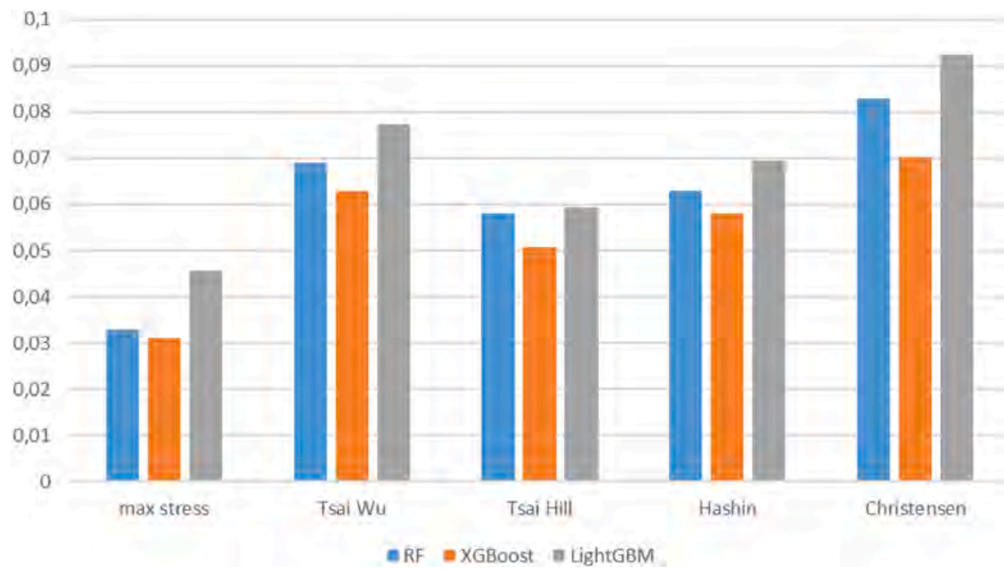


Fig. 19. Comparison of learning methods RMSE performances on critical indexes values.

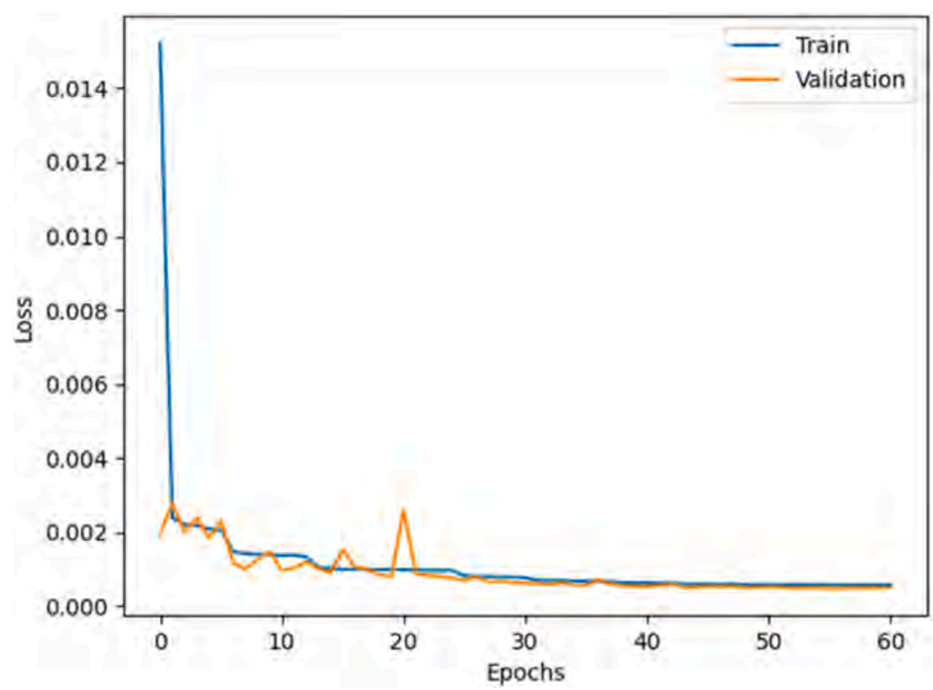


Fig. 20. Max Stress - ANN training vs validation loss function.

However, the scatterplot of predicted versus target values shown in Fig. 21 reveals clear shortcomings in predictive accuracy. As in the Random Forest analysis, the optimal condition would correspond to all points lying on the 45° parity line, indicating perfect agreement between predictions and targets. In contrast, the ANN scatterplot exhibits large deviations from the parity line across both training and test samples. Prediction errors are evident at both low and high FI values, and a saturation phenomenon emerges: the network is unable to predict FI values above a threshold, regardless of the input. This behavior strongly suggests that the ANN failed to learn the underlying functional mapping, resulting in underfitting.

A more detailed error analysis was performed for high-criticality samples (FI > 0.7). As illustrated in Fig. 22, prediction errors tend to increase proportionally with FI magnitude and decrease with sample density. This is consistent with the fact that regions with fewer training instances offer less information for the model to learn robust mappings. Unlike ensemble methods, whose error distributions are more heterogeneous, the ANN errors exhibit a nearly linear trend, further indicating insufficient representational power.

The bar plots in Figs. 23 and 24 summarize error percentages across all FIs for both training and test subsets. The results show nearly identical errors between training and testing sets, highlighting the ANN's inability to generalize. Specifically, the Christensen FI exhibited the largest prediction error, approximately 26% in both training and test sets, while the lowest error was observed for `max_stress`, with values close to 18%. These results confirm that the ANN struggled to fit the data, leading to consistently poor performance across all indices.

From a theoretical perspective, this outcome can be attributed to two main factors. First, the dataset's relatively low variance reduces the

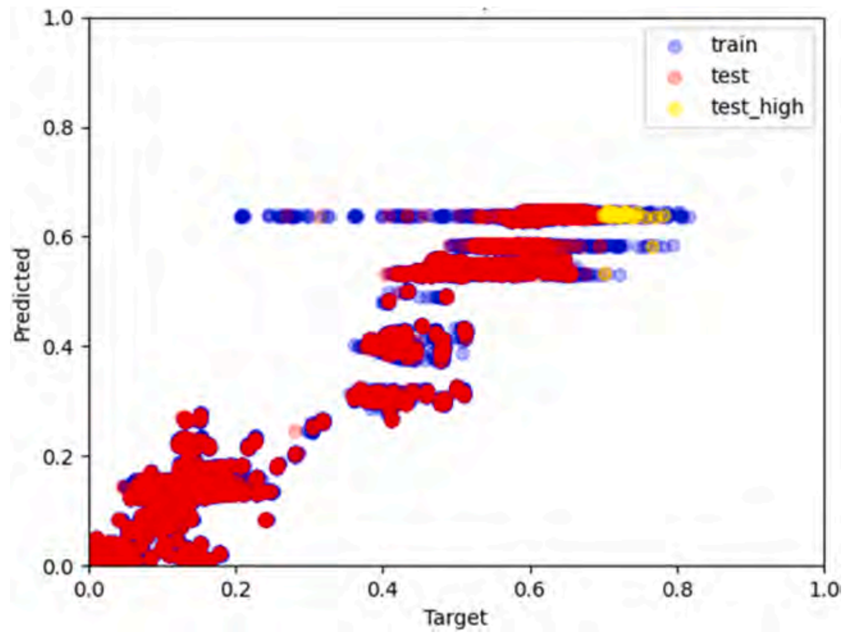


Fig. 21. Scatter plot of Max Stress - ANN prediction versus target.

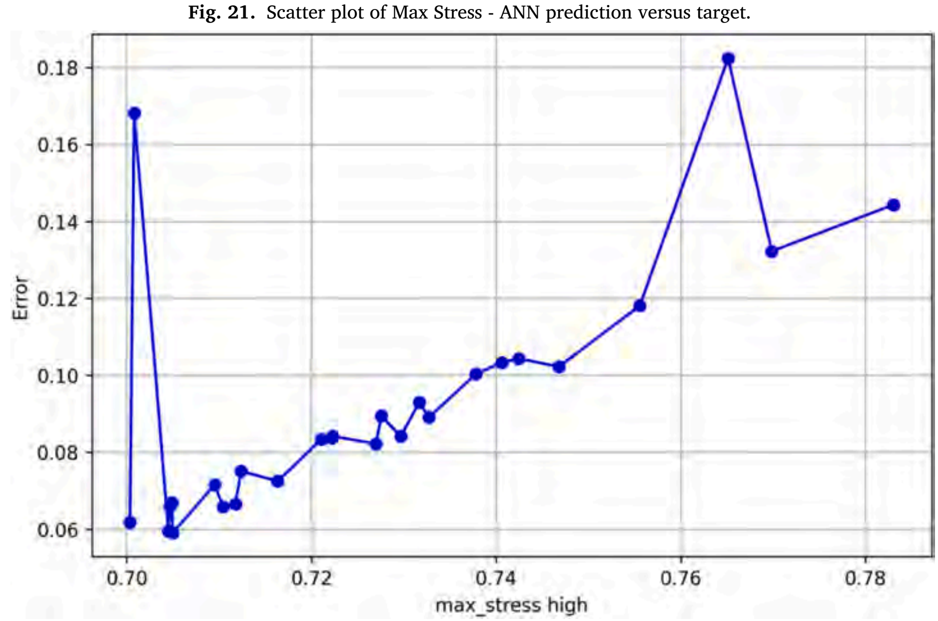


Fig. 22. Max Stress - ANN prediction error on high values.

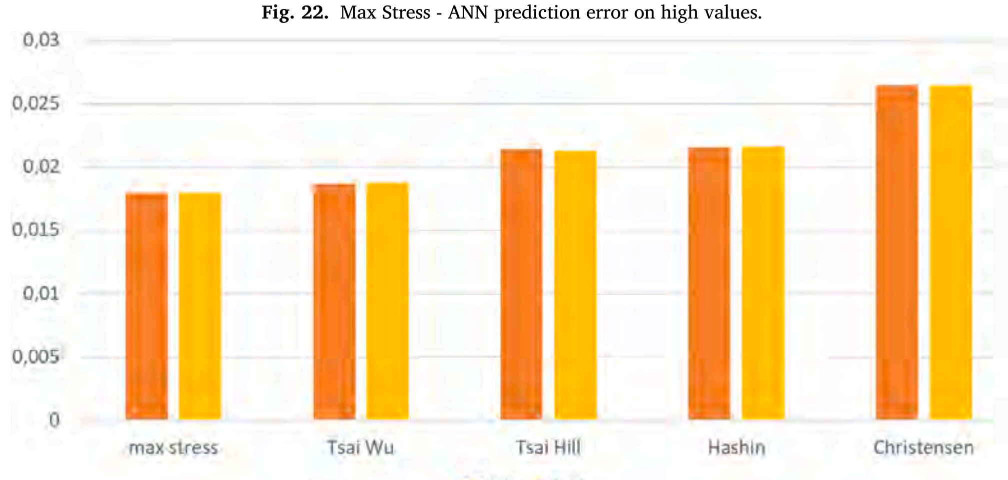


Fig. 23. ANN RMSE performance on all indexes.

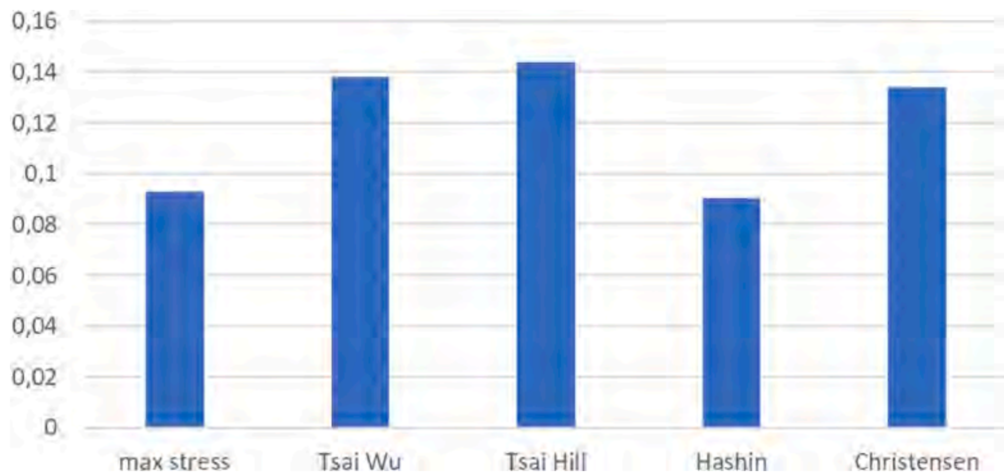


Fig. 24. ANN test over threshold performance on critical indexes values.

Table 6
Cross-validation results for each FI across 5 folds, reporting R^2, MBE, MAE, MSE, and RMSE.

	R^2	MBE	MAE	MSE	RMSE
Results Cross Validation – Max Stress					
Fold 1	0.99987	0	0.00018	0	0.00187
Fold 2	0.99984	0.00001	0.00019	0	0.00209
Fold 3	0.99985	0	0.00019	0	0.00206
Fold 4	0.99986	0	0.00019	0	0.00201
Fold 5	0.99984	0.0001	0.00019	0	0.00210
Mean	0.99985	0	0.00019	0	0.00203
Results Cross Validation – Tsai Wu					
Fold 1	0.99890	0	0.00022	0.00001	0.00346
Fold 2	0.99887	0.00001	0.00021	0.00001	0.00346
Fold 3	0.99880	0.00003	0.00022	0.00001	0.00358
Fold 4	0.99926	0.00001	0.00020	0.00001	0.00279
Fold 5	0.99853	0.00002	0.00024	0.00002	0.00396
Mean	0.99887	0.000014	0.00022	0.000012	0.00345
Results Cross Validation – Tsai Hill					
Fold 1	0.99944	0	0.00014	0.00001	0.00240
Fold 2	0.99946	0.00001	0.00016	0.00001	0.00237
Fold 3	0.99948	0.00001	0.00015	0.00001	0.00232
Fold 4	0.99955	0.00001	0.00015	0	0.00219
Fold 5	0.99928	0.00001	0.00016	0.00001	0.00276
Mean	0.99944	0.000008	0.00015	0.000008	0.00241
Results Cross Validation – Hashin					
Fold 1	0.99899	0.00001	0.00019	0.00001	0.00315
Fold 2	0.99926	0.00001	0.00018	0.00001	0.00270
Fold 3	0.99933	0.00001	0.00018	0.00001	0.00258
Fold 4	0.99907	0.00001	0.00019	0.00001	0.00304
Fold 5	0.99938	0.00001	0.00018	0.00001	0.00249
Mean	0.99921	0.00001	0.00018	0.00001	0.00279
Results Cross Validation – Christensen					
Fold 1	0.99843	0.00001	0.00028	0.00002	0.00399
Fold 2	0.99897	0.00001	0.00027	0.00001	0.00326
Fold 3	0.99885	0.00001	0.00027	0.00001	0.00348
Fold 4	0.99899	0.00003	0.00027	0.00001	0.00323
Fold 5	0.99899	0.00001	0.00026	0.00001	0.00354
Mean	0.99885	0.000014	0.00027	0.000012	0.00350

Table 7
Error metrics (MBE, MAE, MSE, RMSE) computed on high-value FI samples.

Failure Index	MBE	MAE	MSE	RMSE
Max Stress	−0.02110	0.02392	0.00108	0.03287
Tsai Wu	−0.03729	0.04670	0.00477	0.06909
Tsai Hill	−0.03437	0.03773	0.00337	0.05802
Hashin	−0.03768	0.04148	0.00397	0.06298
Christensen	−0.04252	0.05032	0.00677	0.08226

Table 8
RMSE values for different ANN configurations across FIs.

Failure Index	51 features, 3 layers		51 features, 4 layers		11 features, 3 layers	
	Train	Test	Train	Test	Train	Test
Max Stress	0.018	0.018	0.032	0.032	0.048	0.049
Tsai Wu	0.019	0.019	0.023	0.023	0.041	0.041
Tsai Hill	0.021	0.021	0.026	0.026	0.040	0.040
Hashin	0.021	0.022	0.026	0.026	0.036	0.036
Christensen	0.026	0.026	0.029	0.029	0.044	0.044

advantage of employing an artificial neural network, which generally requires high diversity and large sample sizes to exploit its capacity for non-linear function approximation. Second, the imbalance between low- and high-FI cases limits the network's ability to learn patterns associated with extreme structural behaviors, resulting in a systematic underestimation of critical values. These limitations are consistent with the broader literature, which shows that ANNs perform suboptimally on low-variance or imbalanced datasets compared to tree-based ensemble methods.

Table 8 reports the detailed performance metrics across all neural architectures. Overall, the ANN models underperformed compared to ensemble approaches, confirming the incompatibility of neural solutions with the available dataset. While ANNs theoretically offer superior flexibility and scalability, their effectiveness in structural health monitoring tasks strongly depends on the availability of sufficiently diverse training data. Consequently, the ANN results are retained as a purely informative benchmark: they justify the decision to recommend tree-based ML surrogates for the present application rather than neural-network deployment. In the present case, the results highlight the need to either acquire additional high-damage samples or adopt data augmentation strategies before neural networks can become competitive alternatives to ensemble learning.

4. Conclusions

This study proposed a finite-element-driven machine-learning framework for predicting intralaminar failure indices in a damaged composite wing box under static loading. The numerical campaign generated a consistent damage-oriented database in which the affected structural region, damage extent, damage intensity, centroid position, stress state, and five classical composite failure indices were recorded. This setting allowed the learning algorithms to be assessed not only in terms of global prediction error but also with respect to the structural regions and failure-index ranges most relevant to health assessment.

The results indicate that tree-based ensemble regressors are better suited than the tested feed-forward neural-network architectures for the present low-variance and imbalanced tabular dataset. XGBoost and Random Forest achieved the most reliable performance on the hold-out and cross-validation analyses, while the ANN benchmarks showed saturation and underfitting, especially near high failure-index values. The most stable behavior was observed for the Max Stress index, whereas the Christensen criterion remained the most difficult target due to its stronger nonlinear coupling among stress components.

Future work should focus on three limited and technically defined extensions: experimental or mixed experimental-numerical validation on instrumented wing-box components, enrichment of high-index spar-dominated configurations through targeted simulations, and the inclusion of laminate-level descriptors when stacking sequence and delamination effects become part of the modeling scope. Within these limits, the present framework provides a fast screening surrogate that can support preliminary structural health assessment and identify configurations requiring more detailed finite-element or experimental verification.

CRediT authorship contribution statement

Dario Magliacano: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization; **Annalisa Letizia:** Writing – review & editing, Writing – original draft, Conceptualization; **Sara Di Palo:** Validation, Methodology, Investigation, Formal analysis; **Marika Alecci:** Validation, Methodology, Investigation, Formal analysis; **Marco Bazzani:** Supervision, Funding acquisition.

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.

Acknowledgments

The authors acknowledge the support of the Italian Ministry of University and Research (MUR) and the Sustainable Mobility Center (MOST) through the project PNRR - M4C2 - CNMS - Spoke 1, funded under the scheme CN00000023 - PNRR - M4C2 Inv. 1.4 with grant agreement no. 55-PRR22-1155-22-GG002138.

Data availability

Data will be made available on request.

References

[1] F. El Hajj Chehade, R. Younes, Structural reliability software and calculation tools: a review, Innov. Infrastruct. Solut. 5 (1) (2020). https://doi.org/10.1007/s41062-020-0282-4

[2] D. Fanteria, E. Panettieri, A non-linear model for in-plane shear damage and failure of composite laminates, Aerotecnica Missili & Spazio 93 (2014) 17–24. https://doi.org/10.1007/BF03404672

[3] D. Magliacano, High-order vibroacoustic modal analysis framework for fluid-structure coupling, Aerospace 12 (11) (2025). https://doi.org/10.3390/aerospace12110994

[4] M. Filippi, D. Magliacano, M. Petrolo, E. Carrera, Variable-kinematics finite elements for propagation analyses of two-dimensional waveguides, in: 30th AIAA/CEAS Aeroacoustics Conference, 2024, 2024. https://doi.org/10.2514/6.2024-3078

[5] M. Filippi, D. Magliacano, M. Petrolo, E. Carrera, Wave propagation in pre-stressed structures with geometric non-linearities through carrera unified formulation, in: 30th AIAA/CEAS Aeroacoustics Conference, 2024, 2024. https://doi.org/10.2514/6.2024-3028

[6] M. Filippi, D. Magliacano, M. Petrolo, E. Carrera, Wave propagation in prestressed structures with geometric nonlinearities through carrera unified formulation, AIAA J. 63 (1) (2025) 292–308. https://doi.org/10.2514/1.J064695

[7] R.F. Ribeiro Junior, G.F. Gomes, On the use of machine learning for damage assessment in composite structures: a review, Appl. Compos. Mater. 31 (1) (2024) 1–37. https://doi.org/10.1007/s10443-023-10161-5

[8] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, D.J. Inman, A review of vibration-based damage detection in civil structures: from traditional methods to machine learning and deep learning applications, Mech. Syst. Signal Process. 147 (2021) 107077. https://doi.org/10.1016/j.ymssp.2020.107077

[9] A. Tavares, E. Di Lorenzo, B. Cornelis, B. Peeters, W. Desmet, K. Gryllias, Machine learning approaches to damage detection in composite structures combining experimental and simulation domains, Mech. Syst. Signal Process. 215 (2024) 111412. https://doi.org/10.1016/j.ymssp.2024.111412

[10] F. Dipietrangelo, F. Nicassio, G. Scarselli, Structural health monitoring for impact localisation via machine learning, Mech. Syst. Signal Process. 183 (2023) 109621. https://doi.org/10.1016/j.ymssp.2022.109621

[11] L. Lomazzi, R. Junges, M. Giglio, F. Cadini, Unsupervised data-driven method for damage localization using guided waves, Mech. Syst. Signal Process. 208 (2024) 111038. https://doi.org/10.1016/j.ymssp.2023.111038

[12] Q. Wang, H. Qin, L. Jia, Z. Li, G. Zhang, Y. Li, Y. Liu, Failure prediction and optimization for composite pressure vessel combining FEM simulation and machine learning approach, Compos. Struct. 337 (2024). https://doi.org/10.1016/j.compstruct.2024.118099

[13] T. Vrtač, D. Ocepek, M. Česnik, G. Čepon, M. Boltežar, A hybrid modeling strategy for training data generation in machine learning-based structural health monitoring, Mech. Syst. Signal Process. 207 (2024) 110937. https://doi.org/10.1016/j.ymssp.2023.110937

[14] A. Alaimo, A. Barracco, A. Milazzo, C. Orlando, Structural health monitoring procedure for composite structures through the use of artificial neural networks, Aerotec. Missili Spaz. 94 (2015) 14–22. https://doi.org/10.1007/BF03404684

[15] A. Casaburo, D. Magliacano, G. Petrone, F. Franco, S. De Rosa, Optimizing the acoustic properties of a meta-material using machine learning techniques, in: Proceedings of INTER-NOISE 2021 - 2021 International Congress and Exposition of Noise Control Engineering, 2021. https://doi.org/10.3397/IN2021-2294

[16] M.M. Azad, H.S. Kim, Hybrid deep convolutional networks for the autonomous damage diagnosis of laminated composite structures, Compos. Struct. 329 (2024) 117792. https://doi.org/10.1016/j.compstruct.2023.117792

[17] J.M. Ko, Y.Q. Ni, Technology developments in structural health monitoring of large-scale bridges, Eng. Struct. 27 (12) (2005) 1715–1725. https://doi.org/10.1016/j.engstruct.2005.02.021

[18] N. Bakhary, H. Hao, A.J. Deeks, Damage detection using artificial neural network with consideration of uncertainties, Eng. Struct. 29 (11) (2007) 2806–2815. https://doi.org/10.1016/j.engstruct.2007.01.013

[19] D. Magliacano, V. Tufano, A. Letizia, B. Sessa, M. Filippi, Data-driven failure criteria prediction in composite wing boxes using machine learning, Compos. Struct. 363 (2025). https://doi.org/10.1016/j.compstruct.2025.119675

[20] H. Corporation, T700S Standard Modulus Carbon Fiber-Technical Data Sheet, 2025, (Rock West Composites, Datasheet PDF). https://www.rockwestcomposites.com/on/demandware.static/Sites-RWC-Site/Sites-rwc-master-catalog/-/downloads/T700SDataSheet.pdf.

[21] B. Wang, Z. Li, Z. Xu, Z. Sun, K. Tian, Digital twin modeling for structural strength monitoring via transfer learning-based multi-source data fusion, Mech. Syst. Signal Process. 200 (2023) 110625. https://doi.org/10.1016/j.ymssp.2023.110625