POLITECNICO DI TORINO
Repository ISTITUZIONALE

Comparative analysis of ensemble learning techniques for enhanced fatigue life prediction

Original
Comparative analysis of ensemble learning techniques for enhanced fatigue life prediction / Farhadi, S., Tatullo, S., Ferrian, F.. - In: SCIENTIFIC REPORTS. - ISSN 2045-2322. - 15:(2025), pp. 1-18. [10.1038/s41598-024-79476-y]

Availability:
This version is available at: 11583/2998781 since: 2025-04-03T08:45:43Z

Publisher:
Springer Nature

Published
DOI:10.1038/s41598-024-79476-y

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)

13 September 2026

scientific reports

Check for updates

OPEN

Comparative analysis of ensemble learning techniques for enhanced fatigue life prediction

Sasan Farhadi[1]✉, Samuele Tatullo[2] & Francesco Ferrian[1]

Efficient prediction of fatigue life in structural components is crucial for ensuring their integrity and reliability, especially considering the dominant occurrence of fatigue failure in metallic structures within the industrial sectors. Conventional fatigue assessment methods, although theoretically established, are often time-consuming and exhibit limitations due to the intricate nature of the fatigue mechanism. Machine learning models have demonstrated significant potential for enhancing the efficiency of predictions in fatigue life. This research explores the effectiveness of ensemble learning models-boosting, stacking, and bagging-compared to linear regression and K-Nearest Neighbors as benchmarks. Fatigue life prediction is conducted across different notched scenarios using Incremental Energy Release Rate (IERR) measures in addition to the more standard stress/strain field measures. To assess the performance of the proposed models, a comprehensive set of evaluation metrics was performed, including mean square error (MSE), mean squared logarithmic error (MSLE), symmetric mean absolute percentage (SMAPE), and Tweedie score. The findings reveal that ensemble learning models, particularly the ensemble neural networks, stands out as a superior approach for fatigue life cycle assessment compared to other methods. Moreover, the integration of IERR in predicting fatigue life for notched-shape components indicates a promising approach for enhancing the reliability and efficiency of fatigue life predictions in real-world industrial applications.

Keywords Fatigue life cycle, Notched components, Machine learning, Ensemble learning, Incremental energy release rate

Fatigue failures account for a significant proportion of mechanical failures in critical industries such as aerospace, automotive, and heavy machinery. The reliability of these mechanical components under variable load conditions is essential to ensure safety and operational efficiency. Premature failures not only lead to increased maintenance costs but can also result in catastrophic outcomes, emphasizing the importance of accurate fatigue life predictions. Typically, fatigue life assessments have been conducted by experimental and analytical approaches that, while effective to a degree, often fall short when applied to complex real-world applications. These standard methods are challenged by configurations such as notched components, where stress concentrations significantly diminish material fatigue strength. These conventional models, depending heavily on empirical data, struggle with the behavior of materials under varied environmental and loading conditions. The notches, including U-shaped, V-shaped, or more irregular forms such as holes, fillets, and keyseats, act as stress concentrators and affect the integrity of structural components. Predicting the fatigue lifetime in the presence of stress concentrators requires a comprehensive analysis of multiple factors, including geometrical features of the notches, material properties, loading conditions, and environmental influences. Since the initial comprehensive experimental program devoted to the fatigue assessment of metallic components, conducted by Wöhler between 1852 and 1870[1], fatigue experiments and predictions have held a significant role in mechanical design[2]. Focusing on notched components, considering simplified configurations, the peak stress value at the notch edges can be used for accurately predicting the high-cycle fatigue strength of blunt notches. However, this approach tends to be too conservative when applied to sharp notches[3].

Alternative methods have been explored in the last several decades to assess the fatigue lifetime of metallic components. One widely accepted approach, proposed by[4], involves computing the effective stress that damages the fatigue process zone by averaging the stress field near the notch over a characteristic material length. This approach was later simplified by[5], who suggested that the reference stress, relative to the material's fatigue limit, could be determined at a specific distance from the stress raiser-a feature or irregularity in a material or structural

[1]Department of Structural, Geotechnical and Building Engineering, Politecnico di Torino, Torino 10129, Italy. [2]Complex Systems and Networks Program, IMT School for Advanced Studies, Lucca 55100, Italy. ✉email: sasan.farhadi@polito.it

component that causes a localized increase in stress-along the critical crack path. Expanding on this idea, the Theory of Critical Distances (TCD) was introduced as a method for estimating the fatigue strength of notched components in the high-cycle fatigue regime. The accuracy of TCD has been validated for both conventional notched specimens[6–9]and actual components[10,11]. However, TCD approaches may struggle to accurately predict failure loads for notched and cracked components with ligament lengths similar to the critical distance[12,13]. This limitation is addressed by coupled Finite Fracture Mechanics (FFM) approaches, which consider finite crack advancement influenced not only by material properties but also by loading conditions and geometrical characteristics. According to FFM, the stress condition from TCD is coupled with an energy balance involving the incremental energy release rate (IERR). This approach has been applied to assess the fatigue life of various notched geometries under mode I[14–17]and mode III[18] loading conditions.

Although these approaches have proven effective for many geometries, their application is often limited to simplified configurations and loading conditions, requiring complex parameter fitting and finite element analysis for more intricate scenarios. To address these limitations, machine learning (ML) techniques have emerged as a promising and novel approach for modeling fatigue life. The field of ML, pioneered by Arthur Samuel in 1959, has evolved into an extensive and expanding research domain. It encloses diverse methodologies harnessing the computational power of computers to analyze and learn from massive datasets, offering solutions to complex problems. ML models are used in a broad spectrum of research areas, including image segmentation[19,20], natural language processing[21,22], sound event detection[23,24], identification of diseases in healthcare[25,26], environmental science[27,28]mineral exploration and anomaly detection[29–31], and the list could continue given its widespread use today. As a powerful tool, ML versatile applications underscore its key role in shaping the landscape of modern computational methodologies. ML models offer a significant advantage in discerning key features related to fatigue life among different variables.

Recent advancements in ML significantly contributed to fatigue life prediction, addressing a wide array of materials and structural elements. Notably[32], proposed a novel approach using artificial neural networks to create constant life diagrams for metallic materials, improving high-cycle fatigue predictions through empirical data validation. This method set a foundation for subsequent studies including[33]and[34], which applied neural networks to low-cycle fatigue estimation in high-strength steels, demonstrating enhanced accuracy over conventional approaches and highlighting the method's adaptability. ML approaches have been successfully applied to predict fatigue crack propagation in various studies. In the work by[35], neural networks were used to develop predictive models for fatigue life, enabling probabilistic life estimations for damaged structures under stochastically varying input parameters. Similarly[36], combined the Extended Finite Element Method (XFEM) with Radial Basis Function (RBF) neural networks to achieve highly accurate fatigue life predictions. Moreover, the integration of physics-based insights with ML models marked a significant step forward for enhancing prediction performance[37]. and[38]have developed physics-informed machine learning (PIML) frameworks, which not only enhance the estimation of notch fatigue life in aerospace components but also provide a comprehensive analysis of key influential factors. Continuing this trend[39], introduced Physics-Informed Neural Networks (PINNs) that directly integrate defect morphology and physical laws into the training process, significantly enhancing the model's ability to predict fatigue life with high precision. In this context[40], introduced a method based on PINN for fatigue life predictions in defective materials, which takes into account the actual morphology of defects, overcoming the limitations of conventional fracture mechanics approach. The field continues to evolve with innovative methodologies like the fuzzy neuro-inference system introduced by[41]and the use of transformer models by[42] to tackle low-cycle fatigue under thermal-mechanical cycles, indicating the growing complexities and capability of the predictive models.

These developments highlight a trend toward more refined and precise ML methods for predicting fatigue life. Although single models have achieved remarkable results, ensemble models present a promising potential to enhance these predictions by integrating multiple models into a more robust and precise framework[43,44]. Despite the widely study of ensemble methods like Random Forest[45,46], the potential of more advanced approaches, such as ensemble neural networks, remains unexplored specifically for the task of life fatigue predictions. The power of these techniques is to combine many simple "building block" models, which should mitigate the variance and enhance the predictive performance for the specific targets, delivering a robust unified model. To address the mentioned gap and benefiting from the ensemble models as one of the most robust approaches in supervised ML[47], this study conduct an in-depth analysis focusing on materials with different notch shapes under mode-I loading conditions. This research underlines the potential of ensemble models specifically ensemble neural networks for having more reliable and precise fatigue life predictions by offering several considerable contribution to fatigue life prediction. Since to the best of the authors' knowledge, this study represents stress, strain, and IERR for the first time in the context of fatigue life prediction. Furthermore, this research underscores the importance of feature engineering, particularly the use of log-transformation, in enhancing model performance.

The paper is organized as follows: Section Notched components provides a detailed explanation of the characteristics of notched components and the collected data. Applied methods and evaluation metrics are explained in Section Methods. Subsequently, Section Experiments and analysis presents the experiments and analysis. Results and discussions are thoroughly explored in Section Results and discussion. Finally, conclusions are drawn in Section Conclusion, summarizing the key findings.

Notched components

Geometries

This study explores the fatigue life of notched plates under uniaxial tensile load σ_a. The investigation focuses on three distinct configurations: circular holes, U-notches, and V-notches. Specifically, both single and double-edge notches are examined for the latter two geometries.

The notch shapes under consideration herein represent some of the most common and critical features in structural components. These configurations have the potential to significantly influence the reliability and integrity of structures, leading to their sudden collapse. Consequently, the scientific literature has extensively studied the fatigue behavior of notched components through analytical, numerical, and experimental methodologies[48–50]. In the context of simple notched configurations, like those studied in the present study, several methods were developed based on the evaluation of the critical stress and strain values at the notch tip. Alternatively, following the TCD approach, the stress field is evaluated at a critical distance ahead of the stress concentrator apex, along the critical crack path, resting on the assumption of finite crack advancement[7,51]. This characteristic is shared with other nonlocal models, grouping within the framework of FFM, a coupled fracture criterion that allows failure predictions based on the simultaneous satisfaction of a stress condition and energy balance. The latter involves the IERR G_{inc}, derived from the principle of energy conservation before and after crack nucleation over a finite distance. Considering as external load an imposed displacement, under linear elastic assumptions, G_{inc} can be defined as:

$$G_{inc} = \frac{\Delta W_{el}}{L} \tag{1}$$

where L is the crack length and ΔW_{el} denotes the variation of the elastic strain energy due to crack propagation. In the context of these criteria, the IERR, coupled with the stress field, offers a robust framework for predicting the fatigue strength of both blunt and sharp notches[12,15,17,18]. Nevertheless, critical distance-based approaches have different drawbacks, notably complex Finite Element Analysis (FEAs) and parameter fitting. FEAs are conducted to obtain stress, strain, and IERR functions, using bilinear elements under plain strain conditions. The minimum element dimension is determined by a convergence analysis, resulting in a size set to $\frac{r}{100}$ for the configuration with a circular hole and $\frac{d}{100}$ for configurations featuring U and V notches. Stresses, strains and IERR are evaluated along the critical crack path, considering an imposed displacement equal to 1 mm, detailed of FE mesh are depicted in Figure 2. To overcome the computational complexity of these reconstructions, ML techniques emerge as a promising alternative for fatigue life modeling.

The computation of the IERR involves distinct approaches based on the geometric configurations considered. For the circular-hole and double-edge notched configurations (Fig.1a, c and e), considering the symmetry of both the geometry and applied load, symmetric crack propagation is expected to be preferred over an asymmetric one from a theoretical point of view[52]. Conversely, taking into account the single-edge notch, asymmetric crack propagation is supposed to take place from the notch tip along the critical crack path (Fig.1 b and d). In the case of the circular-hole configuration, the stress and strain fields are evaluated along the critical path within a length equal to the radius. To ensure precision in capturing stress and strain gradients near the notch edge, 100 equally spaced points are considered within this zone. Conversely, the IERR function is determined using Eq. 1, obtaining the crack advance by employing a progressive crack unbuttoning approach. Along the same path used for evaluating stress and strain fields, the distance from the notch (l) and the variation of elastic strain energy (ΔW_{el}) due to crack propagation are computed for each node.

Dataset

This study examines a well-balanced dataset containing 188 notched tensile plates. For each plate's failure, the value of stress amplitude (σ_{af}) and the number of cycles to failure (N_f) are considered. Table 1 provided a concise overview of the main features of the dataset. The tests are carried out considering two scenarios: one with a tension-tension setup (loading ratio R=0.1) and another with a tension-compression setup (loading ratio R=−1). It's noteworthy that when R equals 0.1, the stress amplitude fluctuates between 0.1 times σ_{af} and the full value of σ_{af}. Conversely, when R is −1, this range extends from negative σ_{af} to positive σ_{af}. Various material properties, encompassing both steel and aluminum alloys, were considered for this study. The fatigue limit amplitude (σ_0) for the metals falls within the range of 70 to 303 MPa, while the threshold value of the stress

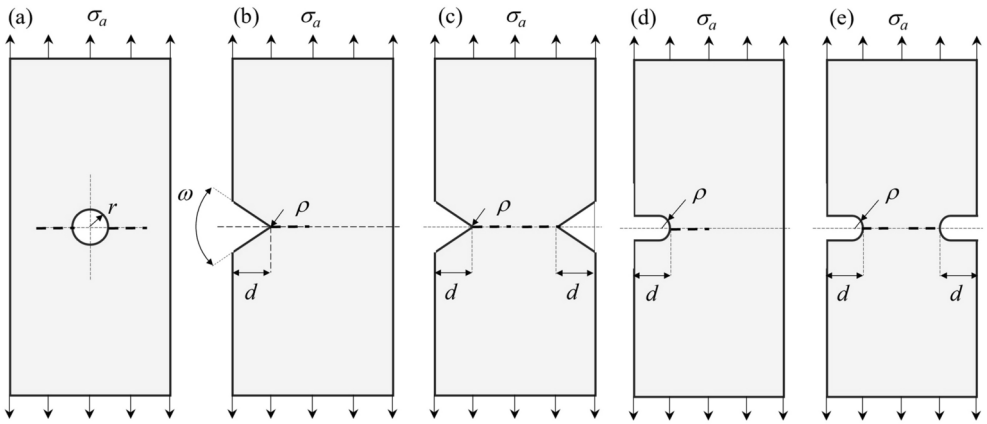

Fig. 1. The geometric configurations: (**a**) circular holes, (**b**, **c**) single and double V-notches, (**d**, **e**) single end double U-notches.

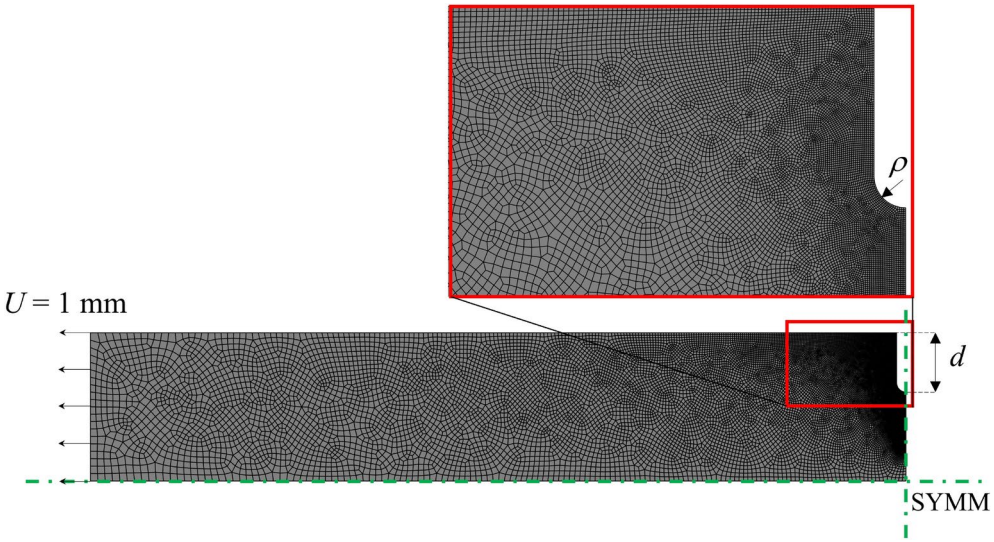

Fig. 2. Mesh and boundary conditions for the FE model of a double U-notched plate.

Type	N	R	E[MPa]	σ_0[MPa]	K_{th}[MPa · m$^{0.5}$]	l_{th}[mm]	Reference
Hole	16	−1	197400	206.3	8.1	1.54	[7]
	16	0.1	197400	163.8	5.9	1.54	[7]
	17	−1	203500	303	6.93	0.52	[53]
	25	−1	72400	124	3.52	0.80	[53]
U-notch	8	−1	197400	206.3	8.1	1.54	[7]
	7	0.1	197400	163.8	5.9	1.30	[7]
	15	0.1	71000	70	2.5	1.28	[54]
	33	0.1	191000	123.5	5	1.64	[54]
V-notch	9	−1	197400	206.3	8.1	1.54	[7]
	17	0.1	197400	163.8	5.9	1.30	[7]
	8	0.1	71000	70	2.5	1.28	[54]
	17	0.1	191000	123.5	5	1.64	[54]

Table 1. Dataset related to notched tensile plates presenting circular holes, U- and V- notches.

intensity factor amplitude (K_{th}) varies between 2.5 and 8.1 MPa·$m^{0.5}$. Moreover, $l_{th} = \frac{K_{th}}{\sigma_0}$ is provided which serves as a fatigue generalization of Irwin's length, a well-known parameter defined within the static framework.

In the circular-hole configuration, various radii r ranging from 0.12 to 4 mm were analyzed. For the U-notched geometry, different notch lengths d were considered, spanning from 5 to 10 mm. Similarly, the V-notched shape configurations contain notches with lengths (d) varying between 4 and 10 mm. Particularly for the V-notched geometry, based on findings by[54], the opening angle (ω) is set at 90°, and the root radius r varies between *0.1*and 0.2 mm. Moreover, referencing data provided by[7], the values adopted are r=0.12 and $\omega = 60°$.

As presented in the previous section, the IERR, Strain, and Stress functions for each geometry were obtained by applying unit displacement. Then, since the analysis is linearly elastic, the functions related to an applied stress equal to σ_{af} are determined considering the reaction force given by the FE model. In this procedure, we need to consider that the stress and strain fields are proportional to σ_{af} whereas the IERR is proportional to the variation of elastic strain energy and, thus, to the square of σ_{af}.

Methods

This section outlines the methodologies employed to predict fatigue life cycles using conventional ML methods and ensemble learning algorithms. The selected models form versatile approaches, addressing the complex patterns in fatigue life prediction scenarios. Subsequent sections detail model explanations and performance evaluation to provide a thorough understanding of the proposed methodologies.

Feature engineering: box-cox transformation

The Box-Cox transformation, introduced by[55], is extensively used to stabilize variance and normalize data distributions. This method is particularly valuable when dealing with non-normal dependent variables in

regression models, improving the robustness of statistical analyses. The transformation is formally defined by the following equation:

$$Y_i = \begin{cases} \frac{X_i^{\lambda}-1}{\lambda} & \text{if } \lambda \neq 0, \\ \log(X_i) & \text{if } \lambda = 0, \end{cases} \tag{2}$$

where X_i represents the observed data and Y_i denotes the transformed outcome. The parameter λ, essential for specifying the nature of the transformation, is generally estimated using maximum likelihood estimation to optimize the fit of the transformed data to a normal distribution. The versatility of the Box-Cox transformation is highlighted by its capacity to adjust to different values of λ, each corresponding to specific types of transformations: $\lambda = 1$ implies no transformation, $\lambda = 0$ leads to a natural logarithm transformation, $\lambda = 0.5$ corresponds to a square root transformation, and $\lambda = -1$ results in an inverse transformation. This adaptability enables the Box-Cox method to correct various levels of skewness within data, enhancing the normality of the distribution.

K-nearest neighbors

The *k*-Nearest Neighbor (KNN) algorithm, introduced by[56], is renowned for its simplicity, operating on the principle that data with similar input entries should share similar values or belong to the same class. The parameter *k* signifies the number of neighbors considered for regression or classification, ranging from 1 to the total training data points. In regression tasks, a crucial choice is whether the weights assigned to data in predictions should depend on their distance, a parameter well-defined within the input space. This distance becomes crucial when determining the *k* nearest neighbors for a new data point *x*, represented by $\{y_i\}_{i=1}^{k}$. These target values contribute to the predictive function, expressed as:

$$f(x) = \sum_{i=1}^{k} \frac{1}{F(d(x_i, x))} y_i \tag{3}$$

here, $F(\cdot)$ represents the distance function, providing flexibility to include a straightforward averaging procedure instead of directly considering distances between inputs and *x* in the computation.

XGboost

XGBoost (eXtreme Gradient Boosting) algorithm belongs to the Gradient Boosting Machine (GBM) algorithms family. Introduced by Schapire[57], boosting is an ensemble learning technique that builds learners sequentially on slightly different datasets (Fig.3). This variation can be achieved by assigning increasing importance to misclassified data in classification or by creating new datasets, e.g., residuals, in regression problems. In GBM, each learner learns from the residuals of the previous one, optimizing the loss function's negative gradient. XGBoost employs tree models as base learners, aiming to minimize the objective function:

$$L(\phi) = \sum_{i} l(\hat{y}_i, y_i) + \sum_{k} \Omega(f_k) \tag{4}$$

where, ϕ represents the ensemble, $L(\cdot)$ is the loss function, $l(\cdot, \cdot)$ is the learners' loss function, $\hat{y}_i$ is the prediction, y_i is the actual value, and Ω is a regularization term preventing overfitting. In XGBoost, Ω takes the form:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2$$

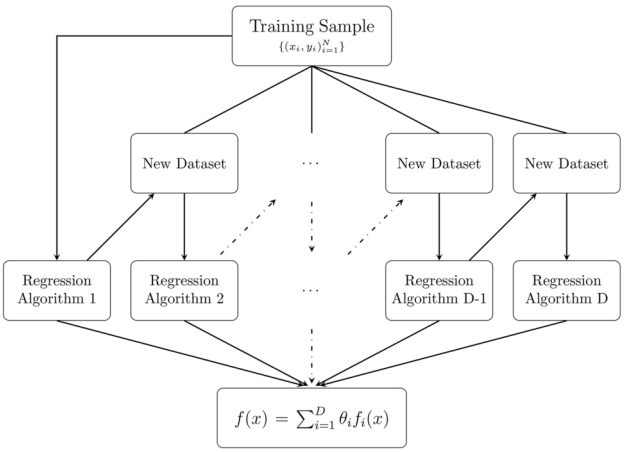

Fig. 3. Ensemble model - boosting.

where T is the number of leafs and w is the leaf weights. Since XGBoost is an additive model, the loss function in the t^{th} iteration becomes:

$$L(\phi^{(t)}) = \sum_i l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t).$$

Stacking

The stacking technique, introduced by Wolpert in[58], revolutionizes ensemble methods by introducing the concept of a "meta-learner." Unlike conventional ensemble methods that use averaging or voting after base learners make predictions, stacking introduces a two-layer learning approach. The first layer, known as the *level-0 model*, consists of diverse base models, each capable of producing different regression or classification results for a given instance. In contrast to other ensemble methods where base learners must be of the same type, stacking encourages diversity among base models. The second layer, *level-1 model*, combines the outputs of the base learners to yield a more accurate and robust regression or classification. For regression tasks, the training dataset is split into two sets: one for training the *level-0 model* using base models and the other for training the *level-1 model*. After training, the base learners perform regression on the second training set, generating a set of potential values for each instance. The *level-1 model* then learns how to effectively combine this information to achieve more accurate predictions. A similar process is applied to classification tasks. Notably, simpler models often yield better results for the *level-1 model*. A visual representation of the stacking process is illustrated in Fig.4.

Random forest

Random Forest (RF), introduced independently by[59]and[60]and further developed by[61], presents a variation of bagged trees. RF construction involves a bootstrapping phase for the training dataset (Fig. 5), creating a random new training set, and a collection of tree-structured learners $f(x, \Theta_k),\ k = 1, \ldots$, where Θ_k are independent identically distributed (*i.i.d.*) random vectors indicating specific quantities that characterize the tree k^{th}. In each split of each tree, m of the p variables are randomly sampled, and the split is allowed to use only one of those m predictors (typically $m \approx \sqrt{p}$). The independence and identical distribution of the $\{\Theta_k\}$ aim to de-correlate the decision trees within the "forest," effectively reducing the ensemble's variance. For regression tasks, the value of a given input x is computed as:

$$f_{\mathrm{rf}}(x) = \frac{1}{D} \sum_{k=1}^{D} f(x, \Theta_k) \tag{5}$$

RF address a potential limitation of bagged trees. If one powerful predictor and many moderately strong predictors exist in the dataset, bagged trees might converge to similar structures. This leads to a situation where the bagged trees closely resemble each other, resulting in highly correlated predictions. Consequently, RF, by introducing randomness in variable selection during each split, acts as a refinement of the bagged-tree approach. According to[61], with an increasing number of trees, the generalized error, denoted as PE^*, converges to a certain value for nearly all sequences $\Theta_1, \ldots$:

$$\mathbb{P}_{X,Y}(\mathbb{P}_\Theta(f(X, \Theta) = Y) - \max_{j \neq Y} \mathbb{P}_\Theta(f(X, \Theta) = j) < 0). \tag{6}$$

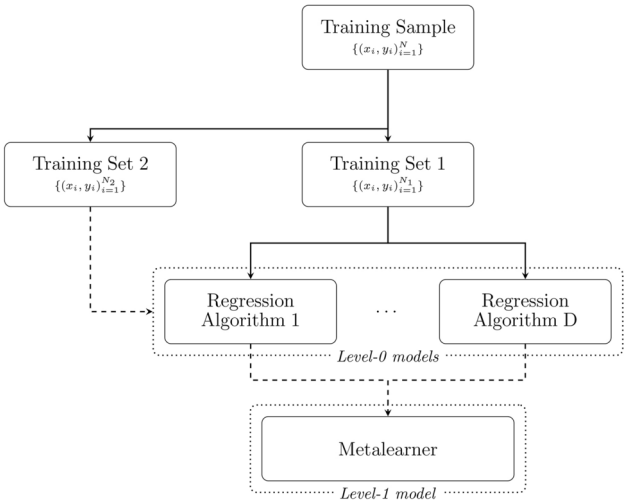

Fig. 4. Ensemble model - stacking.

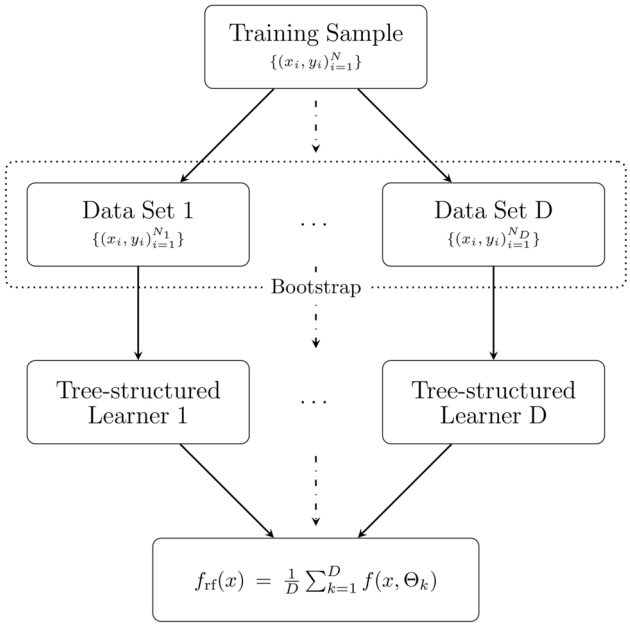

Fig. 5. Ensemble model - random forest.

this equation captures the probability, given a fixed data X and a fixed class Y, concerning all Θ, that the suitable class for X is different from Y. It then computes this probability over all X and Y, representing a general error over all $X, Y,$ and Θ. This characteristic clarifies why RF, as more trees are added, does not overfit but rather converges to a limiting value of the generalization error.

Ensemble neural networks

Ensemble Neural Networks (ENNs), introduced by[62], employ multiple deep learning networks to enhance the predictive accuracy and robustness, particularly effective in complex regression tasks due to their ability to model non-linear data. This ensemble approach utilizes the diversity and combined predictions of diverse deep learning models to address the high variability in conventional neural networks, often arising from their sensitivity to initial setup conditions and data randomness. ENNs can be implemented and designed in several ways[63]. (1) **Initial Condition:** different random weights are initialized to ensure each neural network begins its learning process from a unique starting point. (2) **Network Topologies:** varying architectures by differing numbers of inputs and hidden layers, adding/removing regularization such as dropout,etc. This variation helps in capturing different aspects of the input space comprehensively. (3) **Training Algorithms:** diversifying different algorithms, including stochastic gradient descent and Adam/Nadam optimization. Each algorithm influences the network's learning differently, ensuring the ensemble's ability to explore various errors. (4) **Data Sampling** known as "bootstrap aggregating," this strategy employs different subsets of training data for training individual networks. It not only provides diversity but also ensure varied learning environments, thus enhancing the model's generalization capabilities. To derive outputs, several strategies can be employed, including simple averaging, weighted averaging, voting (commonly used for the classification tasks), and median. In this study, the output of the individual networks, $f_i(x)$ for a given input of x are combined to produce the final output of $F(x)$, define as:

$$F(x) = \sum_{i=1}^{D} w_i \cdot f_i(x), \tag{7}$$

where D represents the number of networks, w_i denotes the weights assigned to each network's prediction, and $\sum_{i=1}^{D} w_i = 1$.

Evaluation metrics

Various metrics were employed to evaluate the performance of the proposed models, including mean squared error (MSE), mean squared logarithmic error (MSLE), Symmetric Mean Absolute Percentage Error (SMAPE), and D^2 Tweedie Score. The MSE is one of the simplest and most common approaches for model evaluation and is defined as:

$$\mathrm{MSE}(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{8}$$

This metric can also be used when there are negative prediction values, and it simply represents the conventional Euclidean distance. In scenarios where the target variable's distribution is skewed or the data exhibit multiplicative relationships, the MSLE offers a finer evaluation compared to MSE. It is defined as follows:

$$\mathrm{MSLE}(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(\log(1+y_i) - \log(1+\hat{y}_i))^2 \tag{9}$$

The logarithmic scaling in MSLE ensures comparability across instances with significantly different magnitudes, contributing to a more robust evaluation. In fatigue life prediction, where the target variable may exhibit characteristics such as skewness and varying scales, MSLE serves as a valuable metric, offering a balanced perspective on prediction performance. In assessing the predictive performance of the models, SMAPE serves as a crucial metric. SMAPE quantifies the prediction performance in terms of percentage errors, offering valuable insights into the proportional difference between predicted and true values. The formula for SMAPE is

$$\mathrm{SMAPE}(y, \hat{y}) = \frac{100}{N}\sum_{i=1}^{N}\frac{2|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}. \tag{10}$$

This metric calculates the mean percentage error for each prediction, providing insights into the magnitude of errors relative to the true values. SMAPE values range between 0% and 200%, with lower values indicating superior predictive performance. This format enhances interpretability, enabling to easily comprehend the magnitude of errors in predictions.

in addition to these primary metrics, the D^2 is utilized as a specialized evaluation metric designed for predictive modeling. It is particularly relevant in scenarios where the target variable demonstrates variance and skewness, making it a suitable choice for fatigue life prediction. The formula for the D^2 Tweedie score is expressed as:

$$d(\hat{y}_i, y) = \begin{cases} (\hat{y}_i - y_i)^2, & \text{for } p = 0 \\ 2 * (y\log(\frac{y_i}{\hat{y}_i}) + \hat{y}_i - y_i), & \text{for } p = 1 \\ 2 * (\log(\frac{y_i}{\hat{y}_i} + \frac{y_i}{\hat{y}_i} - 1), & \text{for } p = 2 \\ 2 * \left(\frac{\max(y_i,0)^{2-p}}{(1-p)(2-p)} - \frac{y_i\hat{y}_i^{1-p}}{1-p} + \frac{\hat{y}_i^{2-p}}{2-p}\right), & \text{else} \end{cases} \tag{11}$$

where, y_i represents the true values from the Tweedie distribution, $\hat{y}_i$ denotes the corresponding predicted values, and p is the Tweedie power parameter. This formula captures the characteristics of the Tweedie distribution, emphasizing the importance of accurately predicting the mean, variance, and skewness of the target variable.

Experiments and analysis

This study evaluates fatigue life prediction across four scenarios-No Field, Stress, Strain, and IERR-using a dataset comprising 188 samples with varied notched geometries, including circular holes, U-notches, and V-notches. The material properties and geometric configurations are deliberately diverse to reflect the range of conditions in real-world structural applications. To analyze the interactions among the physical features, Spearman correlation, a non-parametric measure of rank correlation, was employed, as illustrated by the heatmap in Figure 6a. The primary dataset includes six essential features, including categorical variables representing notch shapes. Through one-hot encoding and the integration of additional field data, the feature set expands to 109, providing a more detailed data representation. The dataset was split into training and testing sets using an 80-20% ratio, a standard approach that balances training effectiveness with evaluation robustness. To ensure the test set accurately reflects the dataset's diversity, stratified sampling was employed, resulting in a balanced composition of 36 samples: 12 from circular holes, 13 from V-notches, and 11 from U-notches, as shown in Figure 6b. This structure helps to mitigate bias and ensures no single class disproportionately affects the model's performance. Importantly, the test set was kept entirely separate from the training data to prevent information leakage, allowing for a robust assessment of the model's generalization ability.

The reliability of ML models are heavily depend on the quality of the input data, making the data prepossessing an essential step in preparing the dataset for effective model training. In this study, a thorough preprocessing phase was applied to ensure the data suitability for fatigue life cycle prediction. As part of the feature engineering process, the Box-Cox transformation technique was considered to correct skewness and stabilize variance across the data, as illustrated in Figure 6c and d. Maximum likelihood estimation was used to determine the optimal lambda (λ) value, which was approximately −0.1, reducing the overdispersion present in the number of cycles variable. Given that this value is close to zero, a log transformation was considered more appropriate, aligning with the literature's support for the Log-Normal assumption in cycle life data. The log transformation compresses large numerical values while expanding smaller ones, making it particularly effective for datasets with heavy-tailed distributions. This approach not only stabilizes variance but also aligns with conventional practices in the field, which assume that the number of cycles before failure follows a Log-Normal distribution. By ensuring more consistent variance, the log transformation contributes to creating a more robust dataset for model training. Furthermore, the dataset was carefully curated to exclude missing values and manage potential outliers. The interquartile range (IQR) method was applied to detect and address outliers, ensuring they did not disrupt the training process or lead to inaccurate predictions. The resulting dataset, enhanced through log transformation

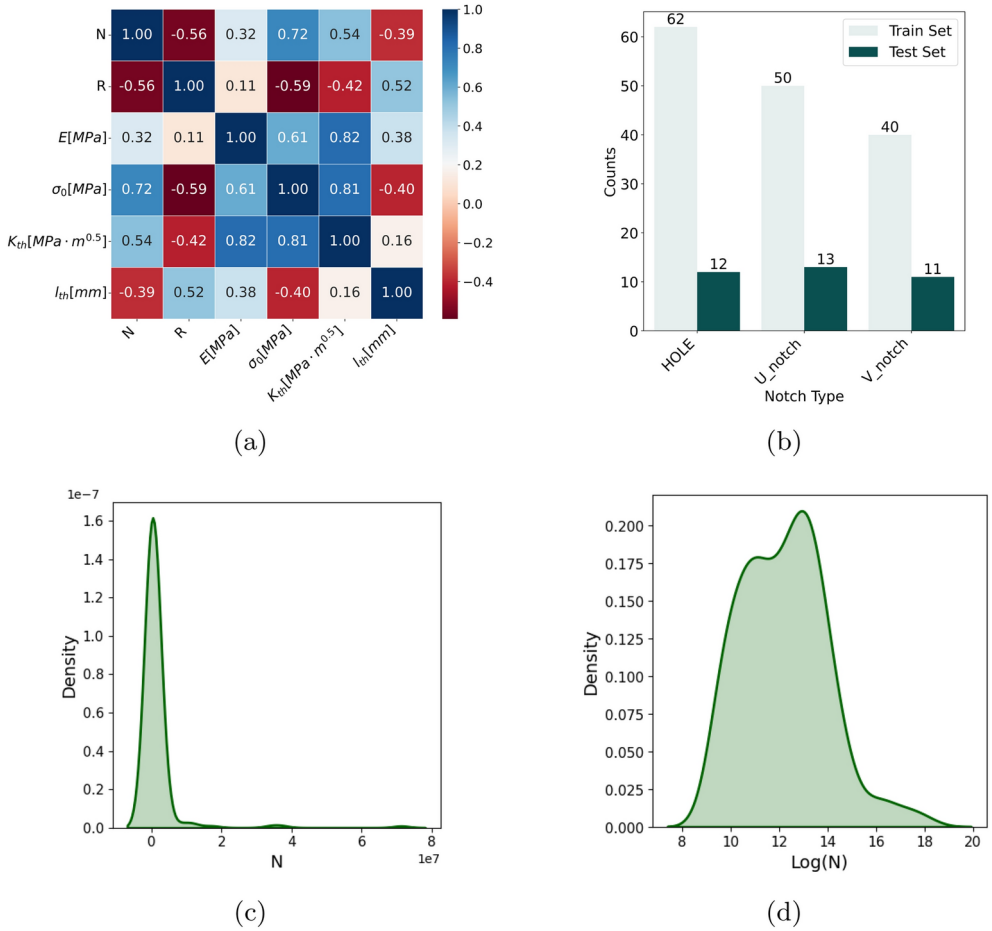

Fig. 6. (**a**) Mechanical properties correlation heatmap, (**b**) Distribution of notch components in train and test sets, (**c**) Actual fatigue life cycles distribution, (**d**) Log-transformed fatigue life cycles distribution.

and outlier handling, provides a solid foundation for model training and prediction. The workflow is illustrated in Figure 7, offering a visual representation of the key steps and processes undertaken in this study.

The careful data preprocessing and feature engineering ensured that the dataset was optimally prepared for use with the proposed ML models. In this study, ensemble models were strategically selected for their ability to handle complex datasets, mitigate collinearity, and enhance predictive performance. XGBoost was implemented as one of the primary ensemble methods due to its robustness and ability to capture intricate data interactions. Stacking further refined predictive performance by integrating outputs from multiple base models, such as decision trees and random forests, through a level-1 decision tree model. RF was used to model the complex data structure inherent to fatigue life cycle predictions, with each ensemble method tailored to the task's unique challenges. Following these well-established ensemble models, ENNs have been utilized to further enhance the prediction task. The preference for ENNs was motivated by their capability and effectiveness in handling non-linear and intricate patterns exhibited in the data. In this study, ENNs combine the strengths of multiple multi-layer perceptrons (MLPs), each trained on bootstrapped samples of the data. This variety of bootstrapped datasets is important for reducing variance and overfitting-common challenges in deep learning applications involving complex physical phenomena. Each MLP in the ENN ensemble was designed with a similar but slightly varied architecture, ensuring that the models offer different perspectives on the same problem, thus improving generalization and enhancing predictions on unseen data. The Nadam optimizer, which combines Adam with Nesterov momentum, was used to accelerate convergence. During training, each model was subjected to its respective bootstrap sample and trained over multiple epochs (set to 100), with early stopping implemented to prevent overfitting. MSE was employed as the loss function to directly target the minimization of prediction errors across fatigue cycle estimations. Furthermore, batch normalization and dropout were employed to further prevent overfitting, and a LeakyReLU activation function was used for all the models. Finally, the outputs of individual networks were aggregated using a weighted averaging scheme, where weights were set based on the validation performance of each network.

Besides, GridSearchCV was employed to determine the optimal hyperparameters in the proposed models. Unlike RandomizedSearchCV, GridSearchCV explores a predefined grid of hyperparameter values, providing a comprehensive search for the optimal configuration. For each hyperparameter, the algorithm evaluates all possible combinations within the specified ranges. Due to its exhaustive nature, GridSearchCV ensures a

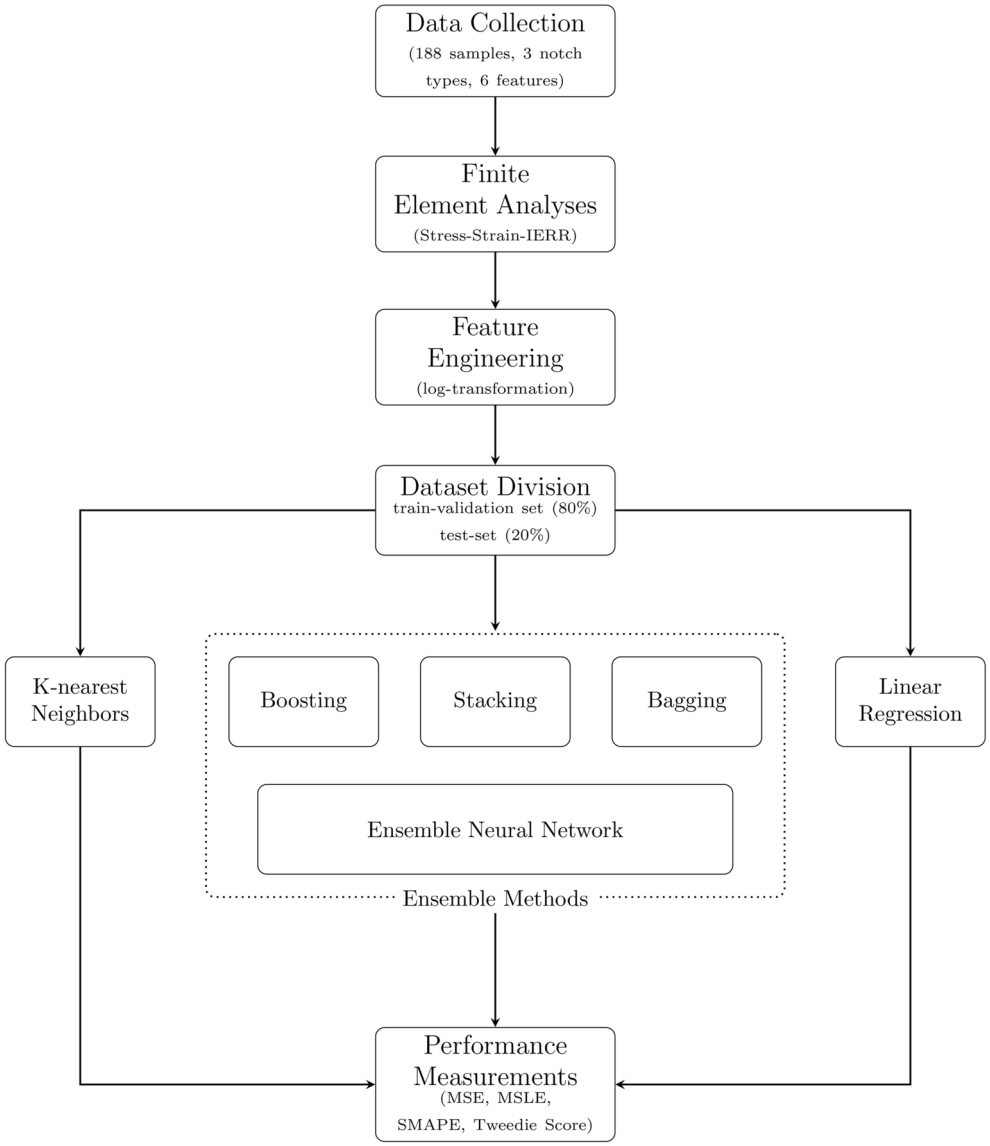

Fig. 7. The strategy workflow to analyze the proposed ML algorithms.

thorough exploration of the hyperparameter space, albeit at the cost of increased computational time. The cross-validation approach stands as an essential approach in optimizing model performance and mitigating overfitting problems. Therefore, in this study, repeated k-fold cross-validation, configured with splits=3 and repeats=2, has been used to enhance the model's robustness and reliability. This method involves partitioning the training set into k-folds, where each fold is considered as a test set, and the remaining $k-1$ folds contribute to model training. This iterative process is repeated multiple times, allowing for a comprehensive evaluation of models under various data splits, ultimately leading to the selection of models with superior performance and optimal hyperparameters. The repeated k-fold cross-validation method offers some advantages over conventional approaches like k-fold or leave-one-out. One of its advantages lies in its capacity to provide a more precise estimate of the model's performance, which is particularly advantageous for small-size datasets. This stems from the repeated process effectively diminishing the inherent variability in model estimation for small datasets. The selected hyperparameters for each model are detailed in Table 2.

Results and discussion

This section evaluates the effectiveness of different ensemble learning models for predicting fatigue life cycles. Building on the methodology outlined in Section 3, the effectiveness of boosting, stacking, bagging, and ENNs is assessed using various metrics, including Mean Squared Error (MSE), Mean Square Logarithmic Error (MSLE), Mean Absolute Percentage Error (SMAPE), and the D^2 Tweedie Score. These metrics provide a comprehensive understanding of each model's strengths and limitations. To perform a robust benchmark, the performance of these ensemble models is compared against two conventional ML approaches, including LR and KNN. LR, a simple yet interpretable linear model, helps clarify relationships between input features

Model	Training complexity	Hyperparameter	Value	Search space
Linear Regression	$O(m^2(n+m))$	alpha	0.5	[0.1, 1]
		penalty	elasticnet	l_1,l_2, elasticnet
KNN	$O(n \cdot m \cdot \log n)$	n_neighbors	15	[5, 15]
		leaf_size	50	[10, 50]
		metric	minkowski	euclidean, manhattan, minkowski
XGboost	$O(T \cdot n \cdot m \cdot \log d_{max})$	n_estimators	50	[10, 50]
		learning_rate	0.8	[0.1, 1]
		max_depth	5	[5, 15]
Stacking	$O(T \cdot n \cdot m \cdot \log d_{RF}$			
level-0 models	$+n \cdot m \cdot \log d_{DT}$			
Decision Tree	$+2n \cdot \log d_{DT})$	max_depth	30	[5, 50]
Random Forest		n_estimators	50	[10, 50]
		max_depth	5	[5, 50]
		criterion	abs._error	squared_error, abs._error
level-1 model				
Decision Tree		max_depth	30	[5, 50]
Random Forest	$O(T \cdot n \cdot m \cdot \log d_{max})$	n_estimators	50	[10, 50]
		max_depth	50	[5, 50]
		criterion	abs._error	squared_error, abs._error
Ensemble NN	$O(T \cdot E \cdot n \cdot \sum_{i=1}^{L} n_i n_{i-1})$	n_layers	20	[5, 30]
		n_neurons	$100^{\dagger}$,10^{*}	[5, 250]
		learning rate	1×10^{-5}	[1e-7, 1e-3]
		activation	leaky_relu	relu, leaky_relu, tanh

Table 2. Selected hyperparameters for proposed models. † =first layer$T$ =No. ensembles/trees$n$ =No. samples$L$ =No. layers$E$ =No. epochs. * =other layersn_i =neurons in layer im =No. features$d$ =max depth.

and the target variable. KNN, a non-parametric algorithm, captures complex relationships without assuming a specific functional form. This comparison offers insights into whether ensemble models deliver improvements over both linear and non-linear approaches. Moreover, as illustrated in Table 2, the computational costs were provided to bring more insights of understanding of each models' characteristics. Although increasing data size can enhance predictive performance, the computational cost of ensemble techniques-where multiple models are trained-can significantly impact efficiency. LR and KNN are the least complex models, whereas ENNs is the most computationally expensive model.However, it should be highlighted that the complexity achieved by the models used in this manuscript are not restrictive since the high computing power of nowadays CPUs. Finally, the analysis of the proposed measurements metrics aims to determine the superiority of the proposed ensemble models over the conventional ML models in predicting life cycle fatigue even if additional computational power is required for ensemble methods.

Performance metrics analysis

The performance of the models, as shown in Table 3, highlights the effectiveness of ENNs compared to conventional models like LR and KNN. In the stress scenario, ENNs emerge as the top performers, achieving the lowest MSLE of 1.99×10^{-3} and the highest Tweedie score of 0.86. XGBoost and RF follow, consistently ranking as the second and third best, underscoring the effectiveness of ensemble models in predicting fatigue life. Notably, KNN slightly outperforms stacking, while LR ranks last with an MSE of 6.89×10^{-1} and a Tweedie score of 0.70. In the strain scenario, ENNs outperform other ensemble models, delivering the lowest MSE (2.38×10^{-1}) and MSLE (1.47×10^{-3}), alongside the highest Tweedie score of 0.90, reflecting their strong capability to capture strain patterns. RF and XGBoost follow, with stacking and KNN showing comparable but lesser performance. In the IERR scenario, ENNs continue to dominate, achieving the lowest MSLE of 7.41×10^{-4}, making it the best-performing model across all scenarios evaluated in this study. XGBoost, RF, and stacking also perform well, outpacing LR and KNN, and further demonstrating the robustness of ensemble methods for predicting fatigue life under varying conditions. However, when only mechanical properties are considered, although ENNs still lead with the highest Tweedie score of 0.86 and the lowest error scores, their dominance is slightly less pronounced, underscoring the importance of feature selection in optimizing model efficacy.

Certain models perform particularly well in specific scenarios. ENNs consistently demonstrate robustness across all scenarios, reinforcing their suitability for fatigue life prediction. In the stress scenario, the diversity offered by ensemble models, particularly ENNs and XGBoost, proves advantageous in handling complex data patterns. In the strain scenario, ENNs show superior performance as well. In scenarios involving IERR and only mechanical properties, although ENNs consistently outperforms other methods, RF indicates as a robust model, suggesting its configuration may be particularly well-suited to the unique data characteristics and challenges present in these scenarios. The observed variations in model performance highlight the importance of tailoring

Scenarios	Model	MSE	MSLE	SMAPE	Tweedie Score
No Fields	Linear Regression	9.90e-01	6.41e-03	7.16e+00	0.55
	KNN	6.63e-01	4.10e-03	4.51e+00	0.72
	XGBoost	4.75e-01	2.89e-03	4.14e+00	0.80
	Stacking	6.65e-01	4.14e-03	5.33e+00	0.71
	Random Forest	3.51e-01	2.08e-03	3.90e+00	0.85
	Ensemble Neural Network	**3.06e-01**	**1.98e-03**	**3.42e+00**	**0.86**
IERR	Linear Regression	8.68e-01	5.47e-03	6.63e+00	0.55
	KNN	5.44e-01	3.39e-03	3.68e+00	0.74
	XGBoost	3.26e-01	1.92e+01	3.2e+00	0.88
	Stacking	3.48e-01	2.07e-03	3.92e+00	0.83
	Random Forest	2.64e-01	1.67e-03	3.35e+00	0.89
	Ensemble Neural Network	**1.41e-01**	**7.41e-04**	**2.25e+00**	**0.95**
Stress	Linear Regression	6.89e-01	4.35e-03	5.26e+00	0.70
	KNN	4.34e-01	2.76e-03	3.03e+00	0.81
	XGBoost	4.22e-01	2.56e-03	3.48e+00	0.83
	Stacking	4.67e-01	2.81e-03	3.85e+00	0.80
	Random Forest	4.07e-01	2.68e-03	3.99e+00	0.82
	Ensemble Neural Network	**3.17e-01**	**1.99e-03**	**3.34e+00**	**0.86**
Strain	Linear Regression	9.90e-01	6.41e-03	7.16e+00	0.55
	KNN	6.61e-01	4.09e-03	4.51e+00	0.74
	XGBoost	5.02e-01	3.28e-03	3.86e+00	0.78
	Stacking	9.36e-01	5.74e-03	6.16e+00	0.60
	Random Forest	3.88e-01	2.50e-03	3.68e+00	0.82
	Ensemble Neural Network	**2.38e-01**	**1.47e-03**	**2.73e+00**	**0.90**

Table 3. Model performance comparison for predicting fatigue life prediction.

ML approaches to the specific characteristics of each scenario. These findings significantly contribute to the ongoing discourse on the optimal selection of ML models for fatigue life prediction across diverse materials and loading conditions.

Conventional methods, such as LR and KNN, often struggle to model the complex, non-linear relationships that exist in fatigue life data. These models rely on assumptions that may not hold in real-world scenarios, particularly when data contains noise or variability, which is common in engineering applications. In contrast, ensemble models, particularly ENNs, can overcome these limitations by effectively capturing non-linear patterns within data, as shown in this research. Moreover, the ensemble nature of ENNs, which combines multiple neural networks, helps reduce and mitigate overfitting and enhances generalization, making them more reliable than conventional methods. Additionally, integrating the IERR within ENNs introduces valuable physical insights often overlooked by conventional approaches. This leads to more robust and realistic fatigue life predictions, especially in cases involving complex geometries or loading conditions. These advantages make ENNs a preferred choice over other models for fatigue life prediction in critical real-world applications where reliability is essential.

Visual representation

Visual representations of the proposed ML models were conducted to provide a better understanding of the predictive outcomes across various scenarios. Figure 8 illustrates the outcomes for the scenario exclusively considering loading conditions and material properties while excluding stress, strain, and IERR field data in the ML model training. In this context, as evident in Figure 8(a), the LR predictions demonstrate inaccuracy, with a significant number of predictions falling outside the ± 2 band, highlighting the challenges the challenges encountered by LR when faced to intricate patterns. As illustrated in Fig. 8 (c-f), it is evident that ensemble models perform significantly better, as they benefit from combining the outputs of multiple learners, which allows them to better manage complex, multi-dimensional relationships and reduce the risk of overfitting. The KNN model (8(b)) closely follows stacking, capturing modes of prediction near the threshold lines and surpassing the LR model, indicating its ability to capture essential patterns, though it still falls short compared to ensemble approaches. These results align well with those presented in Table 3, confirms the consistency and robustness in the evaluation metrics. Incorporating stress, strain, and IERR fields significantly enhanced the model performance, as depicted in Fig. 9, 10, and 11. This comparative analysis underscores the notable superiority of ENNs, over other proposed models in all scenarios. The ENNs show its ability to better capture complexities associated with fatigue life predictions. This is particularly evident in Figure 11(f), where ENNs significantly reduce deviations from actual values, with predictions falling within the ± 1.5 band.

It is essential to emphasize that the deviation of the proposed model's predictions from the defined threshold bands is partially related to the circular hole notch configurations. This observation suggests that U-and V-notch

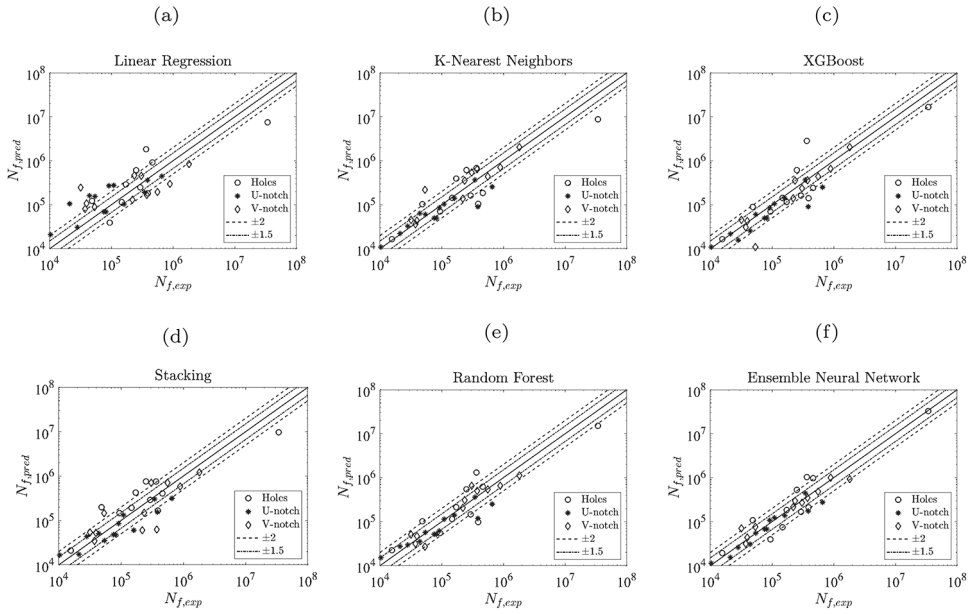

Fig. 8. Comparison of predictive models without field features: (**a**) linear regression, (**b**) K-nearest neighbor, (**c**) XGBoost, (**d**) stacking, (**e**) random forest (**f**) ensemble neural networks.

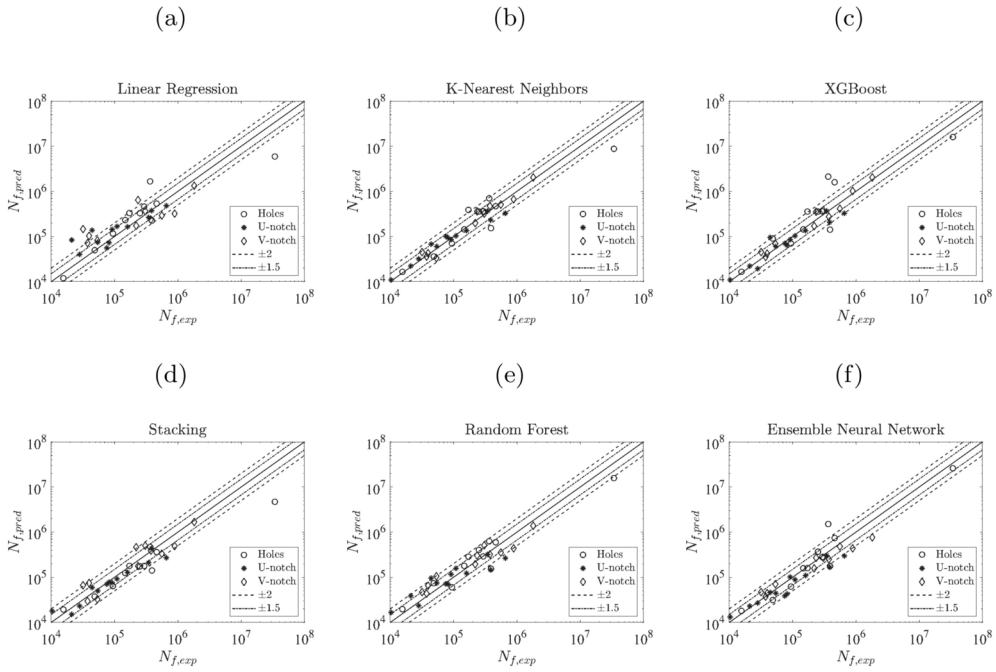

Fig. 9. Comparison of predictive models considering stress scenario: (**a**) linear regression, (**b**) K-nearest neighbor, (**c**) XGBoost, (**d**) stacking, (**e**) random forest (**f**) ensemble neural networks.

configurations are in closer proximity compared to circular hole-notch data, which underscores the importance of incorporating various notch shapes to enhance the models' robustness and generalization ability across different configurations. A specific instance of deviation is notable in the data associated with $N_{f,exp} \simeq 4 \times 10^7$, presenting a challenge for model predictions except ENNs and XGBoost, which exhibits a better prediction. This specifically contributes to the dataset's predominant distribution within the $N_{f,exp}$ range of 10^4 to 10^6, underlining the need for a more balanced representation across the different scenarios for life fatigue prediction to improve the model performance.

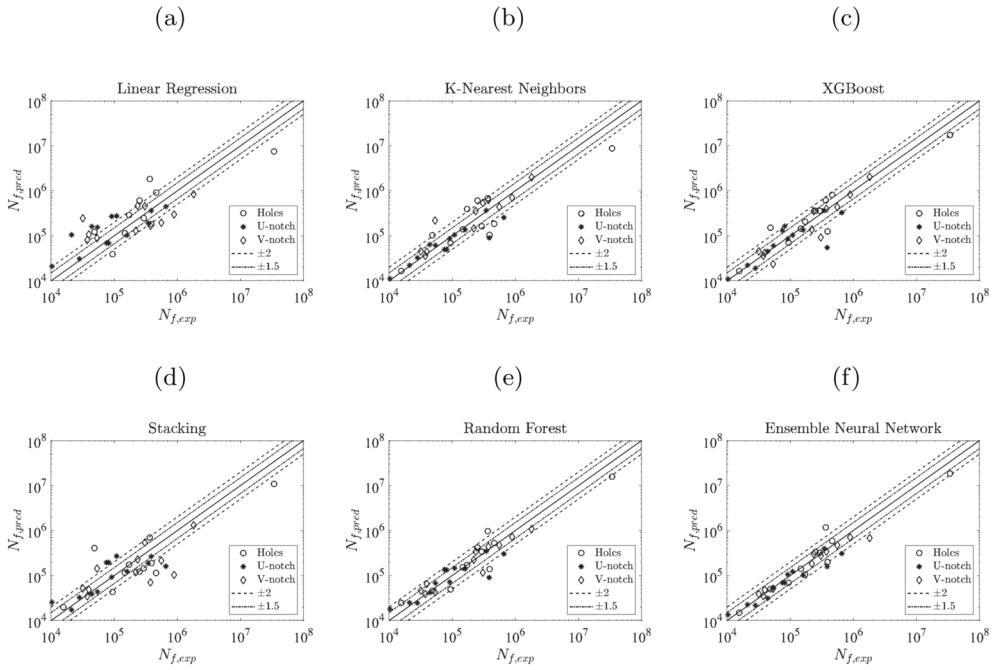

Fig. 10. Comparison of predictive models considering strain scenario: (**a**) linear regression, (**b**) K-nearest neighbor, (**c**) XGBoost, (**d**) stacking, (**e**) random forest (**f**) ensemble neural networks.

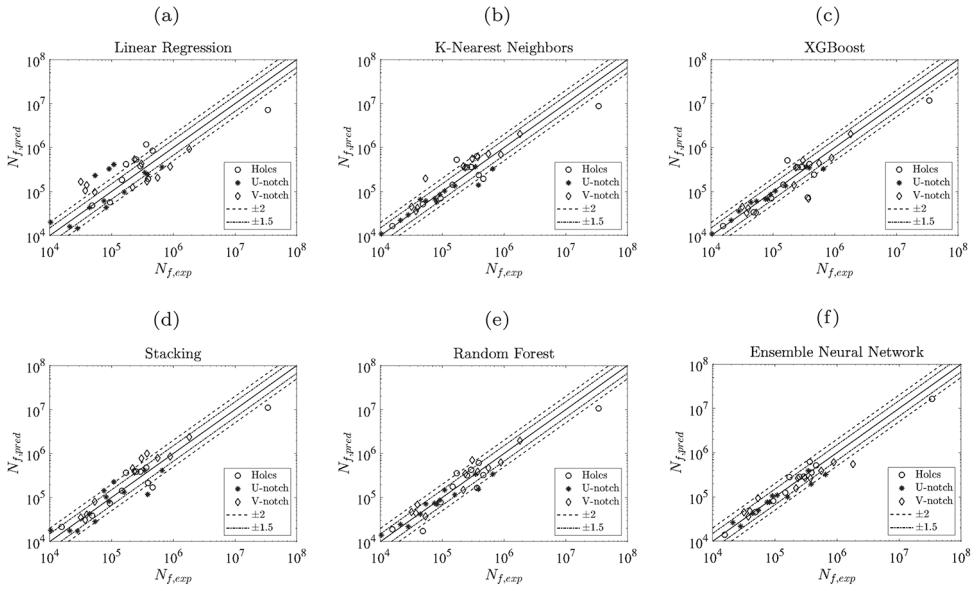

Fig. 11. Comparison of predictive models considering IERR scenario: (**a**) linear regression, (**b**) K-nearest neighbor, (**c**) XGBoost, (**d**) stacking, (**e**) random forest (**f**) ensemble neural networks.

Limitations

Although this study offers valuable insights into predicting fatigue life using various ML techniques, their limitations should be noted. A significant limitation relates to feature selection, which plays a crucial role in capturing the complex patterns of fatigue life cycles. The effectiveness of ML models heavily depends on the quality and relevance of the input features. If essential features are overlooked or inadequately represented, the predictive performance of the models can decline significantly. Additionally, the size and balance of the training dataset are critical factors. In scenarios where the dataset is limited or imbalanced, the models may struggle to learn effectively, leading to biased predictions, particularly in certain fatigue scenarios. The "curse of dimensionality" is another challenge, where an increase in the number of features can actually degrade the

model's performance. As the feature space expands, data points become sparse, making it difficult for models like KNN to find meaningful patterns, which in turn can reduce accuracy.

Ensemble methods, despite their robust predictive capabilities, also have limitations. They often require extensive hyperparameter tuning to perform optimally, which can be time-consuming and computationally expensive. Boosting models, while effective in reducing bias, can sometimes lead to overfitting, resulting in high-variance models. Stacking, which combines multiple models, adds complexity and its success depends on the diversity and quality of the base models. If the base models lack diversity, the ensemble's performance may be compromised. Specifically for ENNs, although they demonstrate strong predictive power, they come with significant computational demands and a risk of overfitting, especially if the architecture is not well-aligned with the data's complexity. Additionally, the process of hyperparameter tuning in ENNs can be complex and time-intensive, often requiring specialized expertise. Another concern is the interpretability of ENNs; their "black box" nature can make it difficult to understand the decision-making process, which is a crucial aspect in many real-world applications. These limitations must be carefully considered when selecting and applying ML models in practice.

Future developments

In the domain of fatigue life prediction, there are several promising routes for future research. One key direction involves expanding the dataset, which could significantly improve the precision of life cycle predictions. By incorporating a broader set of input features, such as grain size, hardness, and other relevant material properties, the quality of model training can be improved, leading to better predictions on unseen data. Furthermore, extending the analysis to include more complex geometries, such as 3D structures or V-notched bars, would provide a more comprehensive evaluation of the model's robustness. This study has focused on mode I fatigue loading; considering mixed-mode states and other real-world loading scenarios could deepen our understanding and lead to more versatile and robust predictive models.

For future expansion of this research, more advanced ML approaches are planned to be explored. Recent advancements in ML, particularly in Physics-Informed Machine Learning (PIML), present a substantial potential for future work. PIML incorporates physical laws directly into the learning process, improving model performance and reducing dependence on large datasets by embedding domain-specific knowledge into the models, as highlighted in recent studies[64,65]. This approach not only enhances predictive accuracy but also improves interpretability, offering a complementary method to purely data-driven techniques. Integrating PIML with advanced neural architectures, such as Graph Neural Networks (GNNs)[66]or Convolutional Neural Networks (CNNs)[67] for spatially distributed data, could further enhance the models' ability to capture complex relationships in fatigue data, especially in the context of irregular geometries and mixed-mode loading scenarios. Therefore, exploring the integration of PIML with ENNs or other advanced architectures could significantly advance the accuracy and reliability of fatigue life predictions, making it a promising direction for future research.

Conclusion

This study conducts a comprehensive exploration of predicting fatigue life cycles through diverse ML methods, ranging from conventional approaches like linear regression and KNN to advanced ensemble learning models, including bagging, boosting, and stacking. The evaluation of these models extended across diverse scenarios, including stress, strain, and IERR, alongside a comparison with a scenario that exclude field features, offering a better understanding of their individual strengths and limitations. The analysis uncovered the limitations of conventional regression models in effectively capturing the intricate patterns exhibited in the prediction of fatigue life. In contrast, ensemble methods, particularly ensemble neural networks, indicate a robust performance across multiple metrics, underscoring their effectiveness in addressing the complexities of this predictive task. The well-trained ensemble models highlight some advantages over the conventional ML techniques, including their proficiency in mitigating overfitting, enhancing generalization, and improving model performance. ENNs emerged as a robust model among the proposed algorithms, consistently excelling in various scenarios. Moreover, this study underscores the significance of careful model selection, emphasizing that a universal model may not exist for all scenarios. In addition to model-specific selection, the impact of feature engineering and feature selection is also taken into account. These factors highlight the complexity of fatigue life prediction and the need for a holistic approach in model development. These findings provide a foundation for future research exploration for model interpretability, feature engineering, and the integration of domain-specific knowledge to enhance model performance in real-world applications.

Data availibility

The data that support the findings of this study are available from the authors on request.

Received: 6 July 2024; Accepted: 11 November 2024
Published online: 01 April 2025

References

1. Collins, J. A. *Failure of Materials in Mechanical Design: Analysis, Prediction, Prevention, 2nd* ed. (Wiley, New York, 1993).
2. Manson, S.S., & Halford, G.R. Fatigue and Durability of Structural Materials. ASM International, Materials Park, Ohio (2006). OCLC: ocm61500113.
3. Smith, R. A. & Miller, K. J. Prediction of fatigue regimes in notched components. *International Journal of Mechanical Sciences* **20**(4), 201–206 (1978).
4. Neuber, H. Theory of notch stresses: principles for exact calculation of strength with reference to structural form and material. *Engineering, Materials Science*, (1958).

5. Peterson, R.E. Notch sensitivity. Metal Fatigue, 293–306 (1959).
6. Susmel, L. & Taylor, D. Fatigue design in the presence of stress concentrations. *The Journal of Strain Analysis for Engineering Design* **38**(5), 443–452 (2003).
7. Susmel, L. & Taylor, D. A novel formulation of the theory of critical distances to estimate lifetime of notched components in the medium-cycle fatigue regime. *Fatigue & Fracture of Engineering Materials & Structures* **30**(7), 567–581 (2007).
8. Wu, Y., Liu, Y., Wang, W., Li, Y., & Geng, R. High cycle fatigue life prediction of single-crystal specimen based on tcd method and crystal plasticity theory. International Journal of Aerospace Engineering, 14 (2023).
9. Li, P., Susmel, L., & Ma, M. The life prediction of notched aluminum alloy specimens after laser shock peening by tcd. International Journal of Fatigue **176** (2023).
10. Taylor, D. Prediction of fatigue failure location on a component using a critical distance method. *International Journal of Fatigue* **22**(9), 735–742 (2000).
11. Cortabitarte, G., Llavori, I., Esnaola, J.A., Blasón, S., Larrañaga, M., Larrañaga, J., Arana, A., & Ulacia, I. Application of the theory of critical distances for fatigue life assessment of spur gears. Theoretical and Applied Fracture Mechanics **128** (2023).
12. Taylor, D., Cornetti, P. & Pugno, N. The fracture mechanics of finite crack extension. *Engineering Fracture Mechanics* **72**(7), 1021–1038 (2005).
13. Gillham, B., Yankin, A., McNamara, F., Tomonto, C., Huang, C., Soete, J., O'Donnell, G., Trimble, D., Yin, S., Taylor, D., & Lupoi, R. Tailoring the theory of critical distances to better assess the combined effect of complex geometries and process-inherent defects during the fatigue assessment of slm ti-6al-4v. International Journal of Fatigue **172** (2023).
14. Liu, Y., Deng, C. & Gong, B. Discussion on equivalence of the theory of critical distances and the coupled stress and energy criterion for fatigue limit prediction of notched specimens. *International Journal of Fatigue* **131**, 105326 (2020).
15. Sapora, A., Cornetti, P., Campagnolo, A. & Meneghetti, G. Fatigue limit: Crack and notch sensitivity by Finite Fracture Mechanics. *Theoretical and Applied Fracture Mechanics* **105**, 102407 (2020).
16. Sapora, A., Cornetti, P., Campagnolo, A. & Meneghetti, G. Mode I fatigue limit of notched structures: A deeper insight into Finite Fracture Mechanics. *International Journal of Fracture* **227**(1), 1–13 (2021).
17. Ferrian, F., Chao Correas, A., Cornetti, P. & Sapora, A. Size effects on spheroidal voids by finite fracture mechanics and application to corrosion pits. *Fatigue & Fracture of Engineering Materials & Structures* **46**(3), 875–885 (2023).
18. Campagnolo, A. & Sapora, A. A FFM analysis on mode III static and fatigue crack initiation from sharp V-notches. *Engineering Fracture Mechanics* **258**, 108063 (2021).
19. Liao, Z., Hu, S., Xie, Y. & Xia, Y. Modeling annotator preference and stochastic annotation error for medical image segmentation. *Medical Image Analysis* **92**, 103028 (2024).
20. Seo, H., Badiei Khuzani, M., Vasudevan, V., Huang, C., Ren, H., Xiao, R., Jia, X., & Xing, L. Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. Medical Physics **47**(5) (2020).
21. Jahan, M. S. & Oussalah, M. A systematic review of hate speech automatic detection using natural language processing. *Neurocomputing* **546**, 126232 (2023).
22. Elizalde, B., Deshmukh, S., Ismail, M. A. & Wang, H. CLAP Learning Audio Concepts from Natural Language Supervision. In *ICASSP 2023–2023 IEEE International Conference on Acoustics* 1–5 (IEEE, Rhodes Island, Greece, 2023).
23. Farhadi, S., Corrado, M., Borla, O. & Ventura, G. Prestressing wire breakage monitoring using sound event detection. *Computer-Aided Civil and Infrastructure Engineering* **39**(2), 186–202 (2024).
24. Mesaros, A., Heittola, T., Virtanen, T. & Plumbley, M. D. Sound Event Detection: A tutorial. *IEEE Signal Processing Magazine* **38**(5), 67–83 (2021).
25. Wei, X., Rao, C., Xiao, X., Chen, L. & Goh, M. Risk assessment of cardiovascular disease based on SOLSSA-CatBoost model. *Expert Systems with Applications* **219**, 119648 (2023).
26. Liu, N. et al. A Novel Ensemble Learning Paradigm for Medical Diagnosis With Imbalanced Data. *IEEE Access* **8**, 171263–171280 (2020).
27. Yuan, Q. et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment* **241**, 111716 (2020).
28. Choubin, B., & Rahmati, O. Groundwater potential mapping using hybridization of simulated annealing and random forest. In: Water Engineering Modeling and Mathematic Tools, pp. 391–403. Elsevier, ??? (2021).
29. Farhadi, S., Tatullo, S., Boveiri Konari, M. & Afzal, P. Evaluating StackingC and ensemble models for enhanced lithological classification in geological mapping. *Journal of Geochemical Exploration* **260**, 107441 (2024).
30. Farhadi, S., Afzal, P., Boveiri Konari, M., Daneshvar Saein, L. & Sadeghi, B. Combination of Machine Learning Algorithms with Concentration-Area Fractal Method for Soil Geochemical Anomaly Detection in Sediment-Hosted Irankuh Pb-Zn Deposit. *Central Iran. Minerals* **12**(6), 689 (2022).
31. Ma, G. & Wu, C. Crack type analysis and damage evaluation of BFRP-repaired pre-damaged concrete cylinders using acoustic emission technique. *Construction and Building Materials* **362**, 129674 (2023).
32. Barbosa, J. F., Correia, J. A. F. O., Júnior, R. C. S. F. & Jesus, A. M. P. D. Fatigue life prediction of metallic materials considering mean stress effects by means of an artificial neural network. *International Journal of Fatigue* **135**, 105527 (2020).
33. Soyer, M. A., Kalaycı, C. B. & Karakaş, Ö. Low-cycle fatigue parameters and fatigue life estimation of high-strength steels with artificial neural networks. *Fatigue & Fracture of Engineering Materials & Structures* **45**(12), 3764–3785 (2022).
34. Wang, Y., Zhu, Z., Sha, A. & Hao, W. Low cycle fatigue life prediction of titanium alloy using genetic algorithm-optimized bp artificial neural network. *International Journal of Fatigue* **172**, 107609 (2023).
35. Giannella, V. et al. Neural networks for fatigue crack propagation predictions in real-time under uncertainty. *Computers & Structures* **288**, 107157 (2023).
36. Wang, M., Feng, S., Incecik, A., Królczyk, G. & Li, Z. Structural fatigue life prediction considering model uncertainties through a novel digital twin-driven approach. *Computer Methods in Applied Mechanics and Engineering* **391**, 114512 (2022).
37. Hao, W. Q. et al. A physics-informed machine learning approach for notch fatigue evaluation of alloys used in aerospace. *International Journal of Fatigue* **170**, 107536 (2023).
38. Wang, L., Zhu, S.-P., Luo, C., Liao, D. & Wang, Q. Physics-guided machine learning frameworks for fatigue life prediction of am materials. *International Journal of Fatigue* **172**, 107658 (2023).
39. Feng, F., Zhu, T., Yang, B., Zhou, S. & Xiao, S. A physics-informed neural network approach for predicting fatigue life of slm 316l stainless steel based on defect features. *International Journal of Fatigue* **188**, 108486 (2024).
40. Salvati, E., Tognan, A., Laurenti, L., Pelegatti, M. & De Bona, F. A defect-based physics-informed machine learning framework for fatigue finite life prediction in additive manufacturing. *Materials & Design* **222**, 111089 (2022).
41. Gao, J., Heng, F., Yuan, Y. & Liu, Y. A novel machine learning method for multiaxial fatigue life prediction: Improved adaptive neuro-fuzzy inference system. *International Journal of Fatigue* **178**, 108007 (2024).
42. Yang, Y. et al. A deep learning approach for low-cycle fatigue life prediction under thermal-mechanical loading based on a novel neural network model. *Engineering Fracture Mechanics* **306**, 110239 (2024).
43. Yang, Y., Lv, H. & Chen, N. A survey on ensemble learning under the era of deep learning. *Artificial Intelligence Review* **56**(6), 5545–5589 (2023).
44. Mian, Z. et al. A literature review of fault diagnosis based on ensemble learning. *Engineering Applications of Artificial Intelligence* **127**, 107357 (2024).

45. Kishino, M. et al. Fatigue life prediction of bending polymer films using random forest. *International Journal of Fatigue* **166**, 107230 (2023).
46. Pałczyński, K., Skibicki, D., Pejkowski, Ł & Andrysiak, T. Application of machine learning methods in multiaxial fatigue life prediction. *Fatigue & Fracture of Engineering Materials & Structures* **46**(2), 416–432 (2023).
47. Ganaie, M. A., Hu, M., Malik, A. K., Tanveer, M. & Suganthan, P. N. Ensemble deep learning: A review. *Engineering Applications of Artificial Intelligence* **115**, 105151 (2022).
48. Landers, C.B., & Hardrath, H.F. Results of axial-load fatigue tests on electropolished 2024-T3 and 7075-T6 aluminum-alloy-sheet specimens with central holes. No. NACA-TN-3631. (1956).
49. Fatemi, A. Fatigue behavior and life predictions of notched specimens made of QT and forged microalloyed steels. *International Journal of Fatigue* **26**(6), 663–672 (2004).
50. Gates, N. & Fatemi, A. Notch deformation and stress gradient effects in multiaxial fatigue. *Theoretical and Applied Fracture Mechanics* **84**, 3–25 (2016).
51. Ezeh, O. H. & Susmel, L. On the notch fatigue strength of additively manufactured polylactide (PLA). *International Journal of Fatigue* **136**, 105583 (2020).
52. Ferrian, F., Cornetti, P., Marsavina, L. & Sapora, A. Finite Fracture Mechanics and Cohesive Crack Model: Size effects through a unified formulation. *Frattura ed Integrità Strutturale* **16**(61), 496–509 (2022).
53. DuQuesnay, D.L., Topper, T.H., & Yu, M.T. The effect of notch radius on the fatigue notch factor and the propagation of short cracks. Mechanical Engineering Publications (1986).
54. Lazzarin, P., Tovo, R. & Meneghetti, G. Fatigue crack initiation and propagation phases near notches in metals with low notch sensitivity. *International Journal of Fatigue* **19**(8–9), 647–657 (1997).
55. Box, G. E. P. & Cox, D. R. An Analysis of Transformations. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **26**(2), 211–243 (1964).
56. Fix, E., & Hodges, J.L. Discriminatory analysis, nonparametric discrimination. USAF School of Aviation Medicine (4) (1951).
57. Schapire, R. E. The strength of weak learnability. *Machine Learning* **5**(2), 197–227 (1990).
58. Wolpert, D. H. Stacked generalization. *Neural Networks* **5**(2), 241–259 (1992).
59. Kam Ho, Tin. Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1, pp. 278–282. IEEE Comput. Soc. Press, Montreal, Que., Canada (1995).
60. Amit, Y., & Geman, D. Randomized Inquiries About Shape: Application to Handwritten Digit Recognition. Chicago Univ IL Dept of Statistics (1994).
61. Breiman, L. *Random forests. Machine learning* **45**(1), 5–32 (2001).
62. Hansen, L. K. & Salamon, P. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**(10), 993–1001 (1990).
63. Jovanovic, R. Z., Sretenovic, A. A. & Zivkovic, B. D. Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings* **94**, 189–199 (2015).
64. Zhu, S.-P., Niu, X., Keshtegar, B., Luo, C. & Bagheri, M. Machine learning-based probabilistic fatigue assessment of turbine bladed disks under multisource uncertainties. *International Journal of Structural Integrity* **14**(6), 1000–1024 (2023).
65. Zhu, S.-P. et al. Physics-informed machine learning and its structural integrity applications: state of the art. *Philosophical Transactions of the Royal Society A* **381**(2260), 20220406 (2023).
66. Zhu, S. et al. High cycle fatigue life prediction of titanium alloys based on a novel deep learning approach. *International Journal of Fatigue* **182**, 108206 (2024).
67. Fei, C.-W. et al. Deep learning-based modeling method for probabilistic lcf life prediction of turbine blisk. *Propulsion and Power Research* **13**(1), 12–25 (2024).

Acknowledgements

The authors would like to thank Prof. Giulio Ventura, Prof. Mauro Corrado, and Prof. Alberto Sapora for their insightful guidance and support throughout the course of this research.

The first author extend sincere appreciation to Dr. Foad Kiakojouri and his assistant, Dorphy Wilendourph, for their invaluable support and guidance throughout the course of this study.

Author contributions

Conceptualization: Sasan Farhadi, Francesco Ferrian Methodology: Sasan Farhadi, Samuele Tatullo, Francesco Ferrian Software: Sasan Farhadi, Samuele Tatullo Validation: Sasan Farhadi, Formal Analysis: Sasan Farhadi, Samuele Tatullo, Francesco Ferrian Investigation: Sasan Farhadi, Francesco Ferrian Resources: Sasan Farhadi, Francesco Ferrian Data Curation: Sasan Farhadi, Samuele Tatullo Writing - Original Draft : Sasan Farhadi, Samuele Tatullo, Francesco Ferrian Writing - Review and Editing: Sasan Farhadi Visualization: Sasan Farhadi, Samuele Tatullo, Francesco Ferrian Supervision: Sasan Farhadi

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025