A defect-based physics-informed machine learning framework for fatigue finite life prediction in additive manufacturing

Salvati, Enrico; Tognan, Alessandro; Laurenti, Luca; Pelegatti, Marco; De Bona, Francesco

DOI
10.1016/j.matdes.2022.111089

Publication date
2022

Document Version
Final published version

Published in
Materials and Design

Citation (APA)
Salvati, E., Tognan, A., Laurenti, L., Pelegatti, M., & De Bona, F. (2022). A defect-based physics-informed machine learning framework for fatigue finite life prediction in additive manufacturing. *Materials and Design*, *222*, Article 111089. https://doi.org/10.1016/j.matdes.2022.111089

Materials & Design 222 (2022) 111089

Contents lists available at ScienceDirect

Materials & Design

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

A defect-based physics-informed machine learning framework for fatigue finite life prediction in additive manufacturing

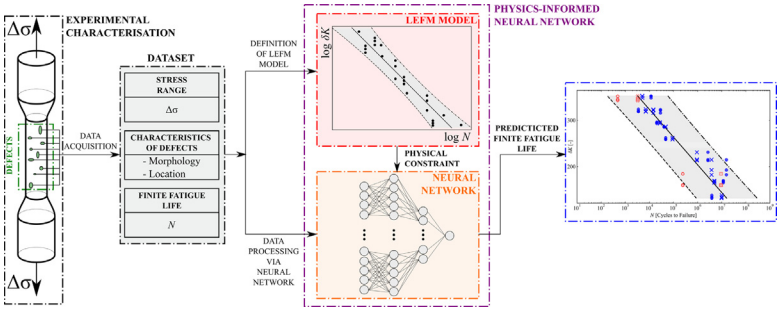
Enrico Salvati [a,*], Alessandro Tognan [a], Luca Laurenti [b], Marco Pelegatti [a], Francesco De Bona [a]

[a] Polytechnic Department of Engineering and Architecture (DPIA), University of Udine, Via delle Scienze 206, Udine, Italy
[b] Delft Centre of System and Control (DCSC), TU Delft University, Mekelweg 2, Delft, the Netherlands

HIGHLIGHTS

- A machine Learning-based approach and phenomenological laws are combined to account for several influencing factors and improve fatigue life prediction.
- A PINN framework is employed to predict finite fatigue life in materials containing defects.
- A newly developed LEFM semi-empirical model is employed to represent the physics of the problem.
- Results from an additively manufactured material are taken from the literature to demonstrate the effectiveness of the proposed method.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 1 June 2022
Revised 2 August 2022
Accepted 21 August 2022
Available online 22 August 2022

Keywords:
Additive manufacturing
Fatigue
Machine learning
Physics-informed neural network
Defects
Fracture mechanics

ABSTRACT

Defects in additively manufactured materials are one of the leading sources of uncertainty in mechanical fatigue. Fracture mechanics concepts are useful to evaluate their influence, nevertheless, these approaches cannot account for the real morphology of defects. Preliminary attempts to exploit a more comprehensive description of defects can be found in the literature, by using Machine Learning. These approaches are notoriously data-hungry and neither physics laws nor phenomenological rules are introduced to assess the soundness of the outcome. Hereby, to overcome this limitation, an approach to predicting fatigue finite life of defective materials, based on a Physics-Informed Neural Network framework, is presented for the first time. The training process of a Neural Network is reinforced by introducing novel Fracture Mechanics constraints. Experimental results obtained from the literature, including detailed defect analysis from computer tomography and fractography, were used to check its accuracy. The proposed predictive tool fully exploits the advanced capabilities of machine learning to account for morphological aspects of defects that could not be accounted for otherwise, while at the same time obeying fracture mechanics laws and requiring a smaller experimental dataset . The approach paves the way for new structural design approaches with an unprecedented degree of accuracy.

* Corresponding author.

https://doi.org/10.1016/j.matdes.2022.111089
0264-1275/

1. Introduction

The development of Additive Manufacturing (AM) techniques has brought an unprecedented degree of geometry flexibility during the design process of mechanical components, which has been enabling the full exploitation of topology optimisation processes [1]. Nevertheless, structural performance – particularly fatigue – of AM materials is still an open issue that needs to be addressed in order to produce reliable engineering components [2]. Besides surface roughness finish, several interplaying microstructural factors rule the fatigue behaviour of these materials, distinctively from those produced by traditional manufacturing processes. Such influencing factors are largely caused by the nature of the process, which consists of a fast-moving heat source that inevitably produces high and inhomogeneous localised cooling speeds. Therefore, the resulting material is greatly influenced by the thermal history and heat fluxes, which are poorly controlled and inevitably give rise to inhomogeneous microstructure. Indeed, AM materials usually show a textured microstructure linked to the heat flux directions, and even crystallographic phase gradients may be present in some cases [3]. Another important consequence of the inefficiently controlled thermal history across the length scales is the origin of residual stress [4]. In the case of AM, residual stresses are present at different length scales, ranging from the intragranular (Type III) to inter-granular (Type II) and eventually macroscopic scale (Type I) [5,6]. All of these stresses play different roles, for instance, Type I residual stress may produce relevant distortions of the manufactured part [7], while Type I and II stresses can importantly alter its fatigue performance. For this reason, a stress relief heat treatment is generally applied [8]. In some other specific applications where superior fatigue performance is sought, post-process mechanical treatments are employed, such as shot/shock peening [9–11], to locally modify residual stress spatial distribution and magnitude.

Although thermal effects can also cause the nucleation of thermal cracks, the predominant presence of crack-like features in AM materials is due to the interplay of localised thermal gradients and dynamics of the material's melt pool producing gas/vapour jets [12] that can promote spatter effects, which are then the principal cause of pore gas formation [13]. Besides gas porosity, whereas insufficient localised heat is provided to the material to be melted, lack-of-fusion defects may arise. Differently from gas porosities, lack-of-fusion pores display a more complex morphology arising from the presence of unmolten particles, oxides, impurities, recrystallised grains and high dislocation density [14]; these characteristic traits make lack-of-fusion pores more detrimental to fatigue resistance.

Despite some post-processing thermo-mechanical treatments have been developed over the recent years to mitigate issues related to the presence of defects (e.g. Hot Isostatic Pressing, HIP) [15], in some cases, these methods either cannot be employed practically or do not produce the expected effectiveness. For this reason, understanding the role of these defects and their influence on the fatigue performance of the material is of paramount importance. Indeed, according to the theory of Fracture Mechanics, small defects may or may not affect fatigue depending on their size, shape, and location in the tested material. In the last decade, Fracture Mechanics concepts have been widely used to assess the fatigue endurance of AM parts considering material defects [16–19], even in the elastoplastic regime [20], thanks also to the original idea of Murakami and his definition of the characteristic size of a defect [21]. Nevertheless, other approaches have also been used [22]. By means of extreme value statistics [23], Romano et al lately employed the statistical distribution of porosity to infer the probability of failure invoking Kitagawa's diagram with El-Haddad's model cast within a Finite Element Methods (FEM) framework [24,25]. Other FEM-based calculation procedures are also available, specifically dealing with local and probabilistic approaches such as P-FAT [26,27] or DARWIN [28].

Detection of defects characteristics in materials can be done through destructive and non-destructive methods. The former refers to material sectioning and optical metallurgical observations, while the latter appeals to a non-invasive technique Computer Tomography (CT) scan [29], for instance. In the specific case of CT, such a technique enables fine morphology details to be captured for a large number of pores, not fully exploited when semi-empirical models - Murakami's for example - are employed. Nonetheless, both experimental and theoretical mechanics has shown how the morphology of initial cracks or defects may affect the actual crack propagation driving force. As analytically demonstrated, concave crack fronts can display increased Stress Intensity Factors (SIFs) [30], meaning that the crack propagation rate is higher in these locations when fatigue loadings are applied externally. Nevertheless, these details are still neglected mainly due to the characteristic crack front evolution that tends to form circular planar cracks, regardless of the initial planar crack profile [31,32]. Additionally, a substantial computational effort is often required to evaluate the number of elapsed cycles required to turn a complex-shaped crack into a circular one, and generally such a number of cycles is negligible as compared with the total number of cycles to failure.

In recent years, the rapid development of Machine Learning-based (ML) techniques has demonstrated a great potential for the recognition of influencing factors and patterns in large datasets; many applications can be found focusing on fatigue problems [11,33–35]. In particular, some approaches have been proposed very recently to deal with the presence of defects in AM metallic materials. For instance, Support Vector Machine (SVM) was successfully applied to infer the fatigue finite life of a selective laser melted Ti-6Al-4V alloy by studying the statistical information obtained by Synchrotron X-ray tomography of the geometric features of the defects [36]. Another application of SVM to deal with defects and associated influencing features was proposed very recently [37]. A further recent research showed an application of a random forest ML model to analyse the finite fatigue life behaviour of a martensitic stainless steel considering the projected defect area and a modified SIF fatigue driving force as inputs to the model [38]. The fatigue life of an AM laser powder bed fused (LPBF) AlSi10Mg alloy was studied using ML as well [39]. In such a particular instance, scanning electron microscopy (SEM) post-mortem examination allowed for the characterisation of the critical defects in order to identify the three most important defect characteristics: projected area, location and morphology. Eventually, ML was employed to understand whether these three characteristics, along with the applied stress magnitude, could achieve a satisfactory prediction of fatigue life. Another ML application to the problem of defect presence was presented by Seunghyun et al. In their research it is shown how CT and surface profilometry characterisation could serve as inputs to a drop-out neural network approach, for Ti64 fatigue bars manufactured via the LPBF technique [40].

It is thus evident how ML techniques can potentially substitute semi-empirical modelling for fatigue prediction when the available data is sufficiently high. On the other hand, semi-empirical models are the only possible choice if a prediction is sought while having only a limited dataset. At the present, one method generally excludes the intervention of the other, thus in some cases, relevant influencing factors are either partially or completely neglected.

The fundamental challenge is therefore the fusion of the two approaches with the aim to reinforce the prediction.

In the last few years, an interesting variant of ML methods has been attracting a great deal of attention, namely Physically-Informed ML [41]; originally developed for Neural Networks (NN) [42]. The distinctive feature of this approach is that it enables to account for physical laws or other phenomenological constraints, e.g. differential equation models, ruling the studied phenomenon to reinforce the prediction. Very recently, such an approach appeared to be effective for the prediction of AM material porosity during manufacturing [43]. Another example is an application of NN to a Ti-6Al-4V AM alloy for the prediction of fatigue life, considering also probability, without considering any material microstructural characteristics such as defects [44]; in this case, the physics-informed approach was referred to as physics-guided. In the field of solid mechanics, the methods was also proposed to identify inhomogeneous elastic properties [45].

In the present work, a novel ML approach – based on the Physics-Informed Neural Network (PINN) framework – is developed and presented to forecast the finite fatigue life additively manufactured metallic materials. This approach enables for taking into account those morphological characteristics of the porosity present in the material that are not contemplated when using classic Linear Elastic Fracture Mechanics (LEFM). By proposing and exploiting the capability of a novel semi-empirical modelling approach to fatigue life based on LEFM, the physics-informed section of the PINN model was employed to reinforce the training process of a properly structured NN. A case-study is eventually shown by analysing a relatively small dataset obtained from the literature regarding an AlSi10Mg alloy produced by selective laser melting, specifically containing material fatigue experimental data and morphological characteristics of the porosity present in the material. Importantly, the analysed samples did not undergo any thermal treatment after fabrication, so the effect of residual stress may be significant. The validation of the devised PINN model is then performed through the so-called K-fold cross validation, i.e. permutations of the samples excluded in the training process of the PINN. To this end, the predicted and actual results in terms of fatigue finite life of independent experimental test samples are compared for each generated fold. In order to prove the benefits offered by the PINN, the training and the validation is replicated using an equivalent NN, i.e. the same PINN where the LEFM physical constraint is deactivated. To this end, both qualitative and quantitative comparison are widely discussed, alongside with a discussion on the implications and potential limitations of the proposed model.

2. Materials and methods

2.1. Experimental dataset

The experimental dataset used in the present work was kindly provided by Romano et al. [17], particularly regarding those which underwent CT analysis. According to the authors, several cylindrical aluminium alloy (AlSi10Mg) samples were additively manufactured by means of an EOS M400® powder-bed machine, which relies on the selective laser melting (SLM) technology. The surface of the samples was turned to remove the intrinsic roughness of the SLM machining process. These samples were fabricated over a period of three years (2015 to 2017) employing the same process parameters, and the diameter of the gauge volume was not the same for all the tested samples: from 4 mm to 6 mm. Additionally, all the samples were fabricated using a layer thickness of 60 μm and a pre-heating of the platform of 200 °C. Throughout the fabrication period, however, the machine's built-in system that recirculates the inert gas and removes the particles was upgraded to improve the quality of the material. Therefore, two different batches of samples were distinguished, namely P1 and P2. Specifically, the samples belonging to P1 and P2 were produced using the original and the upgraded recirculating system, respectively. Furthermore, regardless of the batch involved, two different built orientations were considered, namely vertical (V) and horizontal (H). It is important to highlight that none of the samples was heat treated after the manufacturing process.

Prior to fatigue testing, the samples underwent CT scans in order to reconstruct the accurate morphology and location of hidden, or partially hidden, defects. In this respect, details of the adopted gauge volume can be found elsewhere [17].

In order to characterise the fatigue behaviour of the material, the specimens were subjected to a cyclic load of constant stress amplitude at a load ratio of $R = -1$. In this instance, $\Delta\sigma$ is the nominal externally applied load, whereas N is the corresponding number of cycles to failure. Table 1 provides the outcomes of the fatigue experimental characterisation. After the fatigue testing, the samples were observed through fractography in post-mortem conditions to detect, where identifiable, killer pores, i.e. the pores where the fatigue cracking was triggered. Table 1 reports in which samples this operation turned out to be successful.

The morphology features that are considered pertinent to the present work are: the volume of the defects V_D, the external surface of the defects A_D, and the projection of A_D onto the plane normal to the direction of the applied load. The square root of this projection will be identified as $\sqrt{area}$ in the following sections concerning the LEFM model. According to the features A_D and V_D, the sphericity of the defects is computed as:

$$S = \frac{\pi^{\frac{1}{3}}(6V_D)^{\frac{2}{3}}}{A_D} \tag{1}$$

whereas the equivalent three-dimensional diameter of the defects is given by:

$$d = 2\left(\frac{3V_D}{4\pi}\right)^{\frac{1}{3}} \tag{2}$$

Regarding the location of the defects, the CT scans permitted the distance between the defects and the free surface of the specimen to be assessed. Hereafter, such quantity will be denoted by h.

2.2. Development of a fracture mechanics-based model

2.2.1. Stress intensity factor evaluation

Due to the relatively small size of the defects generally found in AM materials, a LEFM approach is often used to assess the severity of the applied cyclic load concerning the fatigue life. The stress intensity factor, SIF (K), is widely employed to evaluate the crack driving force in LEFM. For fatigue problems, a SIF range (ΔK) is considered as follows:

$$\Delta K = Y\Delta\sigma\sqrt{\pi a} \tag{3}$$

where a is the crack length, while Y is a function depending on the geometry of the problem and loading mode, e.g. crack size, crack morphology, crack position, sample geometry etc.

When dealing with small cracks (up to ~ 1000 μm), compared with the characteristic cross section dimension of the probed material, Murakami demonstrated how even three-dimensional defects can be characterised by simply evaluating the square root of the projected area of the defect over a plane normal to the principal loading direction, $\sqrt{area}$; this characteristic is then considered as the equivalent crack length [21]. Therefore, for small defects, Eq. (3) becomes:

$$\Delta K = Y\Delta\sigma\sqrt{\pi\sqrt{area}} \tag{4}$$

Table 1
List of tested samples and fatigue details [17]. As concerns the built orientation, H and V stand for horizontal and vertical, respectively.

Progressive Number, i	Batch, P#	Built orientation	Stress range, $\Delta\sigma$ [MPa]	Experimental fatigue life, N [number of cycles]	Killer defect detectable?
1	1	H	400	474	YES
2	1	H	220	2,622,640	YES
3	1	H	360	3432	YES
4	1	V	200	15,242,310	YES
5	1	V	180	11,352,768	YES
6	1	V	295	46,255	YES
7	1	H	298.5	19,806	NO
8	1	V	301.6	39,538	NO
9	1	V	180	237,485	NO
10	1	V	378.3	11,465	NO
11	1	V	156	3,795,336	NO
12	2	H	374.4	28,201	NO

As widely reported by Murakami [21], Y turns out to be dependent predominantly on the distance between the centre of the defect and the sample free surface – for short cracks. In particular, Murakami proposed the following empirical condition to discriminate (sub-)surface defects and embedded cracks:

$$h/r < 1.25 \tag{5}$$

where h is the distance from the sample free surface of the defect centre, and r is the radius of the equivalent circular planar defect defined as:

$$r = \sqrt{\frac{area}{\pi}} \tag{6}$$

Therefore, according to LEFM, the crack coefficient turns out to be: $Y = 0.65$ for (sub-)surface defects and $Y = 0.5$ for embedded (or bulk) cracks.

$$Y = \begin{cases} 0.65 & h/r < 1.25 \\ 0.50 & h/r \geq 1.25 \end{cases} \tag{7}$$

Particular case of sub-surface defects.

For a more precise evaluation of the ΔK, the equivalent crack length for sub-surface defects should be corrected to account for the fraction of material laying between the defect and the free surface that does sufficiently restrain the defect to open when the load is applied [21,46]. In order to do so, an *effective area* ($area_{eff}$) is generally employed. The concept of the effective area stems from the relatively fast propagation of sub-surface cracks, preferentially towards the sample free surface, along with those directions where the crack driving force is predominant. In this scenario, fatigue crack growth occurs rapidly to form a secondary defect geometrical configuration ($area_{eff}$) and these fatigue cycles allowing for a crack to grow from $area$ to $area_{eff}$ are neglected; this is a generally accepted assumption [46].

Given the difficulty in achieving an accurate estimate of $area_{eff}$ - especially when dealing with three-dimensional defects - the exact shape of the defect can be conveniently and efficiently replaced by accounting for the distance of the sample free surface h to the centre of the equivalent circular defect again by a planar circular crack with the equivalent radius, r, according to Fig. 1. It is important to report that the centre of the equivalent circular defect is coincident with the centroid of the defect itself.

By using these dimensions, an approximate value of the $area_{eff}$ can be defined through the following equation:

$$area_{eff} = \frac{\pi r^2}{2} + 2hr = area\left(\frac{1}{2} + \frac{2}{\pi}\frac{h}{r}\right) \tag{8}$$

In this way, the projected defect area is increased by maximum 1.3 times at the boundary between bulk and sub-surface defects ($h/r = 1.25$), and this multiplying factor turns out to be 1 when $h/r = \pi/4$ (limit of applicability of the present sub-surface defect).

By substituting $area_{eff}$ in Eq. (4), and assuming Y = 0.65 as per surface defects, the following expression can be written:

$$\Delta K^{sub-surf} = 0.65\left(\frac{1}{2} + \frac{2}{\pi}\frac{h}{r}\right)^{1/4}\Delta\sigma\sqrt{\pi\sqrt{area}} \tag{9}$$

As it can be seen, Eq. (9) maintains the same structure of Eq. (4). Therefore, a modified version of Y can be defined straightforwardly as:

$$Y^{sub-surf} = 0.65\left(\frac{1}{2} + \frac{2}{\pi}\frac{h}{r}\right)^{1/4} \tag{10}$$

which accounts for the aforementioned defect surface increment. It is worth outlining that this handy estimation procedure might lead to either an overestimation or underestimation of the effective area depending upon the shape of the defect. In any case, this effect is considered to be negligible in the present paper.

Surface cracks and generalised correction in SIF calculation.

The correction on the effective area decays when the boundaries of internal defects meet the material sample free surface. In this scenario, there is no need to consider an effective area of the defect, but the actual projected area ($area$) can be used. As quickly mentioned earlier, as it can be seen in Eq.(8), a further criterion must be implemented to discriminate this class of cracks, thus:

$$h/r \leq \pi/4 \tag{11}$$

In summary, a modified version of Y, namely Y_{eff} is readily defined:

$$Y_{eff} = \begin{cases} 0.65 & \text{if } h/r \leq \pi/4 \\ 0.65\left(\frac{1}{2} + \frac{2}{\pi}\frac{h}{r}\right)^{1/4} & \text{if } \pi/4 < h/r < 1.25 \\ 0.5 & \text{if } h/r \geq 1.25 \end{cases} \tag{12}$$

Hence, the SIF range (ΔK) can be promptly evaluated through equation (4), where Y is now replaced by Y_{eff}:

$$\Delta K = Y_{eff}\Delta\sigma\sqrt{\pi\sqrt{area}} \tag{13}$$

2.2.2. Normalised fatigue driving force

Recently, it was shown how accounting for the presence of defects in tested materials using LEFM concepts can reduce the scatter of fatigue life data. In other words, the classic $S-N$ (a commonly used denomination for the $\Delta\sigma - N$ curve) curve can be effectively replaced with a $\Delta K - N$ with the advantage, in some

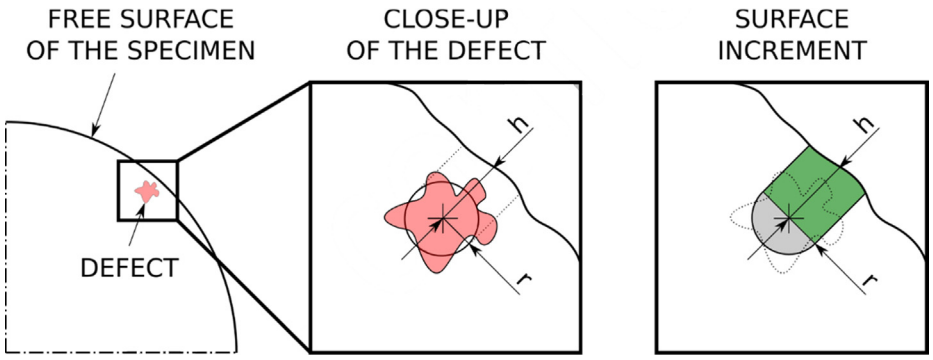


Fig. 1. Graphical illustration of equivalent planar circular crack and its surface increment for sub-surface defects. The projected defect area is highlighted in red, and the radius of its area equivalent circular domain is r. The distance of the equivalent domain to the free surface is h. The effective area of the defect is approximatively given by the sum of half of the equivalent circular defect (in grey) and the portion of material connecting to the free surface (in green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cases, of having a less scattered reference design curve [47–49]. When considering additively manufactured materials containing small cracks, the SIF range can be considered as the driving force which governs the relation between the applied stress range and the number of cycles to failure.

In a recent publication, Murakami showed how the normalisation of $S-N$ curves using the fatigue limit $\Delta\sigma/\Delta\sigma_w$ can be effective to describe the finite fatigue life behaviour of different batches of samples belonging to the same parent material; different batches were referred to different sizes of defects [50,51], where $\Delta\sigma_w$ is the material fatigue limit. In the present paper, an analogous normalisation is proposed to deal with short cracks, with the goal of reducing the scatter band of the master fatigue life curve. To this end, the normalisation is done by using the fatigue SIF threshold, ΔK_{th}. Therefore, $\Delta K/\Delta K_{th}$ becomes the normalised fatigue driving force. However, such normalisation is not straightforward given that in the short crack region ΔK_{th} is strongly affected by the initial crack size as widely reported by Murakami and Ritchie [21,52]. Specifically, this result is attributed to plastic deformations and crack closure effects [53–55]. In the present work, the normalisation was done by evaluating ΔK_{th} according to the empirical law proposed by Murakami and Endo [56].

Murakami and Endo showed that ΔK_{th} can be easily found with remarkable accuracy for many metallic materials, and defect sizes ranging from 10 μm to 1000 μm, through the following equation in which only the material Vickers hardness (HV) is used as the intrinsic material property:

$$\Delta K_{th} = C \times 10^{-3}(HV + 120)\left(\sqrt{area}\right)^{1/3} \tag{14}$$

where C is a constant that depends on the defect position, i.e. 2.77 for bulk cracks and 3.30 for sub-surface or surface cracks [21], following the same distinction discussed earlier. It is noteworthy that the units of ΔK_{th} and $\sqrt{area}$ are MPa $m^{1/2}$ and μm, respectively.

At this point, in order to fully exploit Murakami's relation, it is useful to normalise ΔK in the $\Delta K - N$ also with respect to the material intrinsic mechanical property HV and other remaining constants, therefore Eq. (14) can be rearranged as:

$$\frac{\Delta K_{th}}{C \times 10^{-3}(HV + 120)} = \left(\sqrt{area}\right)^{1/3} \tag{15}$$

Thus, the normalisation is done by dividing Eq. (13) as follows:

$$\frac{\Delta K}{\Delta K_{th}} C(HV + 120) = \frac{Y_{eff}\Delta\sigma\sqrt{\pi\sqrt{area}}}{\left(\sqrt{area}\right)^{1/3}} \tag{16}$$

It is worth noting that the constant 10^{-3} in Eq. (15) vanishes if it is decided to express $\sqrt{area}$ in μm. Given that the coefficient C varies according to the same rule of Y_{eff} (Eq. (12)), it is possible to include the functional relationship of both C and Y_{eff} into a single function Y^*:

$$Y^* = \begin{cases} 0.1970 & \text{if } h/r \le \pi/4 \\ 0.1970\left(\frac{1}{2} + \frac{2}{\pi}\frac{h}{r}\right)^{1/4} & \text{if } \pi/4 < h/r < 1.25 \\ 0.1805 & \text{if } h/r \ge 1.25 \end{cases} \tag{17}$$

Eventually, the driving force of the fatigue life can be expressed as:

$$\frac{\Delta K}{\Delta K_{th}}(HV + 120) = \frac{Y^*\Delta\sigma\sqrt{\pi\sqrt{area}}}{\left(\sqrt{area}\right)^{1/3}} = \delta K \tag{18}$$

This parameter will be often recalled within the present paper as the *normalised SIF range*: δK.

2.2.3. Fatigue curve regression and prediction band

As mentioned earlier, recent research has shown how scattered fatigue data could be fit by a regression line by employing the SIF range [47–49,57], similarly to that classically done using Basquin's law for $S-N$ diagrams. According to the E739 ASTM standard [58], the number of cycles to failure N is assumed to be the dependent variable, whereas the independent variable is the normalised SIF range. Therefore, the fitting power function in the present work is expressed as:

$$N = A\ \delta K^B \tag{19}$$

where A and B are fitting coefficients that can be found through the Ordinary Least Squares method (OLS) after taking the logarithm of both sides:

$$\log N = \log A + B \log \delta K \tag{20}$$

It is worth mentioning that ASTM's standard also reports the protocol to compute the simultaneous *confidence* interval for the regression model, which evaluates the confidence interval of the sole regression line. Given that the introduction of the empirical formulation of fatigue life in the present paper is aimed at providing an interval – that will be later discussed and implemented within the PINN calculation scheme – in which the fatigue failure is expected to lay in, a so-called non simultaneous *two-sided prediction interval* approach is fully exploited [59]. This method is particularly useful since it provides prediction intervals interpreted as a band which includes, with a given confidence, a single future observation. This interval takes into account both the random vari-

ability of a future observation and the epistemic variability of the regression estimates. As a particular case, if the sample size tends to infinity, the prediction interval converges to the scatter band of the sole regression.

According to the two-sided prediction model, the prediction band can be represented by adding and subtracting the predictive term to Eq. (20) as follows:

$$\log N = \log A + B\log \delta K \pm s\Delta P \tag{21}$$

where s is the square root of the estimator for the variance associated with $\log N$:

$$s = \frac{1}{n-2}\sum_{i=1}^{n}\left[\log N_i - (\log A + B\log \delta K_i)\right]^2 \tag{22}$$

and ΔP is the semi-amplitude of the prediction interval, which is computed through:

$$\Delta P = t(n-2,\alpha/2)\sqrt{1+\frac{1}{n}+\frac{\left(\log \delta K - \overline{\log \delta K}\right)^2}{\sum_{i=1}^{n}\left(\log \delta K_i - \overline{\log \delta K}\right)^2}} \tag{23}$$

where $t(n-2,\alpha/2)$ is the $\alpha/2$ quantile of the t-distribution with $n-2$ degrees of freedom, n is the sample size, and $\overline{\log \delta K}$ is the sample mean.

2.3. Physics-Informed Neural Networks

2.3.1. Introduction to the Physics-Informed Neural network framework

Physical sciences heavily rely on physical and empirical models to describe observed phenomena. In the field of solid mechanics, many phenomena are modelled by partial differential equations (PDE), e.g. equilibrium and conservation laws, or empirical and semi-empirical formulations, often of a power law nature, a few examples: fatigue fracture mechanics [55,60,61], fatigue life [62], and creep [63]. A sufficiently accurate description of the physical problem using these equations must ensure that the experimental observation employed for the calibration of their constants is not affected by noise and, additionally, that other variables of the problem that cannot be modelled are thought to be negligible.

Particularly for problems dealing with PDE, over the past decades, considerable effort has been devoted to the development of numerical methods to solve them. Nevertheless, with high-dimensional systems, this approach can quickly become computationally infeasible. On the other hand, semi-empirical formulations may neglect important factors which hamper their accuracy. For these reasons, PINNs have been developed and honed to enforce the mere numerical modelling inherited from NN with the underlying physics of the observed phenomenon. PINN-based methods have turned out to be particularly powerful to facilitate and strengthen the training process in case of a small data regime [41]. The working principle of PINNs is illustrated in Fig. 2. Since the presented approach does not involve PDE-driven models, but considers semi-empirical formulations (see Section 3.1), it is useful to illustrate a generalised PINN framework with a generic model F, which provides the output $\boldsymbol{y} \in \mathbb{R}^V$ according to the input $\boldsymbol{x} \in \mathbb{R}^U$ and $t \in \mathbb{R}$. This can be mathematically stated as:

$$\boldsymbol{y} = F(\boldsymbol{x}, t) \tag{24}$$

where y, x and t could represent any pertinent quantities of interest involved when constructing the model. Based upon a suitable NN, the goal of the PINN approach is to build a surrogate model of F, namely $\boldsymbol{F^{NN}}$:

$$\boldsymbol{y^{NN}} = F^{\boldsymbol{NN}}(\boldsymbol{x}, t, \boldsymbol{\theta}) \tag{25}$$

Where $\boldsymbol{y^{NN}} \in \mathbb{R}^V$ is the output predicted by the NN, and $\boldsymbol{\theta} \in \mathbb{R}^W$ is the vector containing the parameters of the NN.

The parameters of the NN are determined in order to minimise the error of the predictions with respect to the training dataset while satisfying the model in Eq. (24). To this end, the inputs $\boldsymbol{x}$ and t are elaborated by the NN which eventually provides the corresponding predicted output $\boldsymbol{y^{NN}}$. Let $\boldsymbol{y}_e \in \mathbb{R}^V$ be the expected output associated with $\boldsymbol{x}$. At this point, $\boldsymbol{y^{NN}}$ and $\boldsymbol{y}_e$ are compared, and the loss $L^{\boldsymbol{NN}}$ (usually a mean squared error) is computed accordingly. With reference to Fig. 2, $\boldsymbol{y^{NN}}$ is also numerically processed and fed into the model. It should be mentioned that $\boldsymbol{y^{NN}}$ could only satisfy the model (Eq. (24)) to a certain degree, thus resulting in a residual which, in turn, represents the additional loss $L^{\boldsymbol{M}}$. Subsequently, the total loss L is given by an appropriate combination, usually a weighted sum, of $L^{\boldsymbol{NN}}$ and $L^{\boldsymbol{M}}$:

$$L = L^{\boldsymbol{NN}} \oplus L^{\boldsymbol{M}} \tag{26}$$

This combination is concisely represented by the symbol $\oplus$. Finally, L is minimised by updating the parameters of the NN via backpropagation. It is important to note that $L^{\boldsymbol{M}}$ is the portion of the total loss that guarantees the fulfilment of underlying prescribed physics.

2.3.2. PINN for finite fatigue life prediction

In the present study, the PINN framework is used to enforce a NN with phenomenological constraints, referred to as *LEFM fatigue model*, so that the physics of the problem is preserved. Aiming at illustrating the pursued calculation algorithm approach, the training process of the NN is comprehensively illustrated in Fig. 3.

With reference to Fig. 3, a set of input parameters can be identified and labelled as *experimental datasets*, which contains all the experimental information obtained from fatigue tests ($\Delta\boldsymbol{\sigma}_j$, $\mathbf{N}_j$), CT analysis ($\sqrt{\boldsymbol{area}_j}$, $\mathbf{h}_j$, $\mathbf{S}_j$, $\mathbf{d}_j$) and, if available, fractography measurements of the killer defect ($\sqrt{\boldsymbol{area}_j^*}$, $\mathbf{h}_j^*$), for each tested sample j ($j = 1, 2\cdots, m$); where m is the number of fatigue tested samples.

As far as the NN branch is concerned, the sole feature information obtained from CT scans of each tested sample was used since the prediction is meant to be done before the sample is experimentally tested and thus the actual killer defect cannot be known beforehand. Therefore, a filtering operation is executed to prepare the relevant data which will be used as inputs to the NN. Such a filter relies on the identification of a number of defects (l) that gives rise to the largest magnitudes of δK and following sorting in ascending order, i.e. $\delta K_1 < \delta K_2 < \ldots < \delta K_l$. The size of l (also referred to as the number of potential killer defects) is chosen by the user – the case study reported in the next section will show how an appropriate value of l can be judiciously selected. Therefore, all the defect features associated with the filtered killers are used as inputs into the NN's, the whole set of inputs can be summarised as:

$$\begin{aligned}
\sqrt{area}_j &= [\,\sqrt{area}_{1,j} \quad \sqrt{area}_{2,j} \quad \cdots \quad \sqrt{area}_{l-1,j} \quad \sqrt{area}_{l,j}\,] \\
h_j &= [\,h_{1,j} \quad h_{2,j} \quad \cdots \quad h_{l-1,j} \quad h_{l,j}\,] \\
S_j &= [\,S_{1,j} \quad S_{2,j} \quad \cdots \quad S_{l-1,j} \quad S_{l,j}\,] \\
d_j &= [\,d_{1,j} \quad d_{2,j} \quad \cdots \quad d_{l-1,j} \quad d_{l,j}\,]
\end{aligned} \tag{27}$$

where j denotes the index of the j-th sample. Additionally, each tested sample j is characterised by its fatigue testing conditions $\Delta\sigma$ and its resulting number of cycles to failure N, so that this information can be succinctly gathered into the following vectors:

$$\begin{aligned}
\Delta\sigma &= [\Delta\sigma_1 \quad \Delta\sigma_2 \quad \cdots \quad \Delta\sigma_m\,] \\
N &= [N_1 \quad N_2 \quad \cdots \quad N_m\,]
\end{aligned} \tag{28}$$

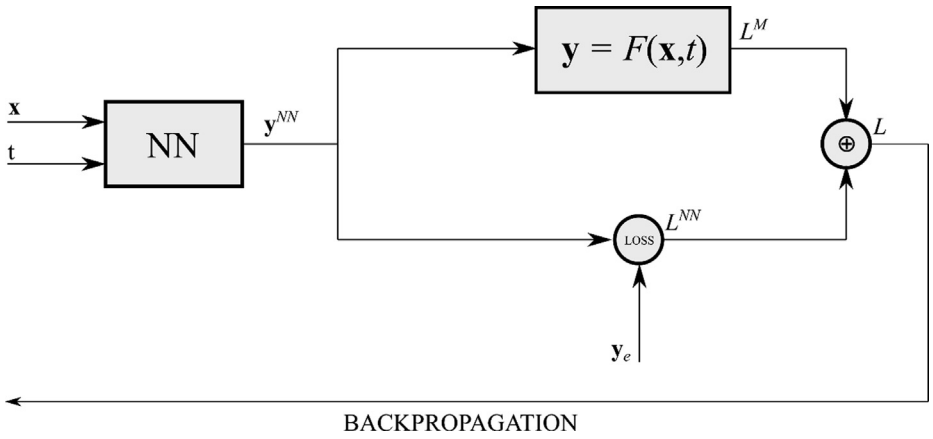


Fig. 2. Conceptual representation of PINN. The input x and t are elaborated by the NN which estimates the output $\boldsymbol{y}^{\boldsymbol{NN}}$. Then $L^{\boldsymbol{NN}}$ is computed according to $\boldsymbol{y}^{\boldsymbol{NN}}$ and the expected output $\boldsymbol{y}_{\boldsymbol{e}}$. Simultaneously, $\boldsymbol{y}^{\boldsymbol{NN}}$ is processed through the model $\boldsymbol{y} = F(\boldsymbol{x}, t)$ giving rise to an additional loss $L^{\boldsymbol{M}}$. Finally, $L^{\boldsymbol{NN}}$ and $L^{\boldsymbol{M}}$ are combined and backpropagated.

Importantly, the data in Eq. (27) and Eq. (28) should be pre-processed in order to prepare the input vector which will be fed into the NN. With reference to the j-th sample, its associated input vector $\boldsymbol{x}_j$ can be constructed in a block-partitioned form:

$$\boldsymbol{x}_j = \left[\sqrt{\boldsymbol{area}_j}|\mathbf{h}_j|\boldsymbol{S}_j|\boldsymbol{d}_j|\Delta\sigma_j\right] \in \mathbb{R}^{\boldsymbol{u}\times\boldsymbol{l}+1} \tag{29}$$

Thereby, the dimension of the input layer of the NN (n. of input neurons) is $u \times l + 1$, where u is the number of defect characteristics (or features) used in this paper ($u = 4$ in this instance: $\sqrt{\boldsymbol{area}_j}$, $\mathbf{h}_j$, $\mathbf{S}_j$ and $\mathbf{d}_j$) and 1 refers to the applied load $\Delta\sigma_j$. Besides, N_j is the expected experimental output associated with $\boldsymbol{x}_j$. Once defined the architecture of the NN, its predicted output in terms of the number of cycles to failure will be given by $N_j^{NN} = \boldsymbol{F}^{\boldsymbol{NN}}(\boldsymbol{x}_j = [\sqrt{\boldsymbol{area}_j}|\mathbf{h}_j|\boldsymbol{S}_j|\boldsymbol{d}_j|\Delta\sigma_j], \boldsymbol{\theta})$ in agreement with Eq. (25).

At this point, the mean squared error between the predicted N_j^{NN} and the actual number of cycles to failure N_j is computed by taking the logarithms, thus providing the loss L_j^{NN}:

$$L_j^{NN} = \left(\log N_j^{NN} - \log N_j\right)^2 \tag{30}$$

Such a loss is summed to an additional loss derived from the physics side L_j^M, which will be defined afterwards, giving rise to the total loss L_j.

As widely discussed earlier, this model requires knowledge of the sole N, $\Delta\sigma$, h and $\sqrt{area}$ to build the normalised SIF range vs. number of cycles to failure diagram ($\delta K - N$) and its associated regression and prediction band (framed with red dashed lines in Fig. 3). To define this model both CT and fractographic measurements can be employed, giving the priority to the actual killer defect information obtained from fractography, while, if this information is missing, then a number of l potential killer defects are exploited.

Given the logarithm of the first n.l normalised SIFs of the j-th sample, namely $\log\delta K_{j,1}, \log\delta K_{j,2}, \cdots, \log\delta K_{j,l}$, the $\delta K - N$ model (see Eq. (20)) can be used to predict the logarithm of the associated potential killers expected finite fatigue lives, namely $\log N^*_{j,1}, \log N^*_{j,2}, \cdots, \log N^*_{j,l}$. Reasonably, a loss function which penalises the difference between each $\log N_{j,i}$ ($i = 1, 2, \cdots, l$) and $\log N_j^{NN}$ (predicted by the NN) should be sought. The definition of such a loss function exploits an important assumption of the model: $\log N$ for a fixed level of $\log\delta K$ follows a Normal Distribution [58]:

$$\log N \sim \mathrm{Nor}(\log N^*, s) \tag{31}$$

where $\log N^*$ is given by Eq. (20) evaluated at $\log\delta K$, and s is the variance (Eq. (22)). Upon normalising the random variable $\log N$, Eq. (31) provides:

$$\log N \sim \frac{1}{\sqrt{2\pi}s} e^{-\frac{1}{2}\left(\frac{\log N - \log N^*}{s}\right)^2} \tag{32}$$

In view of defining a suitable loss function it is useful to normalise the values provided by Eq. (32) so that this function lays in $[0, 1]$. Hence the previous equation can be restated as:

$$\log N \sim e^{-\frac{1}{2}\left(\frac{\log N - \log N^*}{s}\right)^2} \tag{33}$$

Given that such a normal distribution describes what the probability of occurrence of N is expected at a given δK, the aim is to exploit this phenomenological condition into the definition of the loss function for the physics constrains of the PINN.

Considering Eq. (33), one could recognise the squared difference between the expected fatigue life $\log N^*$ and its potential prediction $\log N$, in the argument of the exponential function. This squared difference resembles a loss function, similar to what is implemented in NN back propagation process. Nevertheless, this form cannot be used as a loss function to enforce the physics constrains since the presence of the exponential penalises the predictions of $\log N_j$ when approaching the expected value $\log N^*$. In the present work it is proposed to revert the behaviour of the function displayed in Eq. (33) such that it can be effectively used as a loss function. The fundamental idea is to have a function that provides a null loss function at the mean value of the normal distribution and a loss function approaching a unitary value far away from the mean value.

To this end, the general form of the loss function driving the physics branch of the PINN framework is devised as follows:

$$L^M = 1 - e^{-\frac{1}{2}\left(\frac{\log N - \log N^*}{s}\right)^2} \tag{34}$$

This equation can be readily adapted to handle the prediction given by the NN, $\log N_j^{NN}$, and the expected value $\log N^*_{j,i}$ provided by the $\delta K - N$ model, as far as the j-th sample is concerned:

$$L_j^M = 1 - e^{-\frac{1}{2}\left(\frac{\log N_j^{NN} - \log N^*_{j,i}}{s}\right)^2} \tag{35}$$

In order to fully define the loss function which pertains to the physical part of the PINN of the present paper, Eq. (35) was opportunely generalised for the first n.l normalised SIFs:

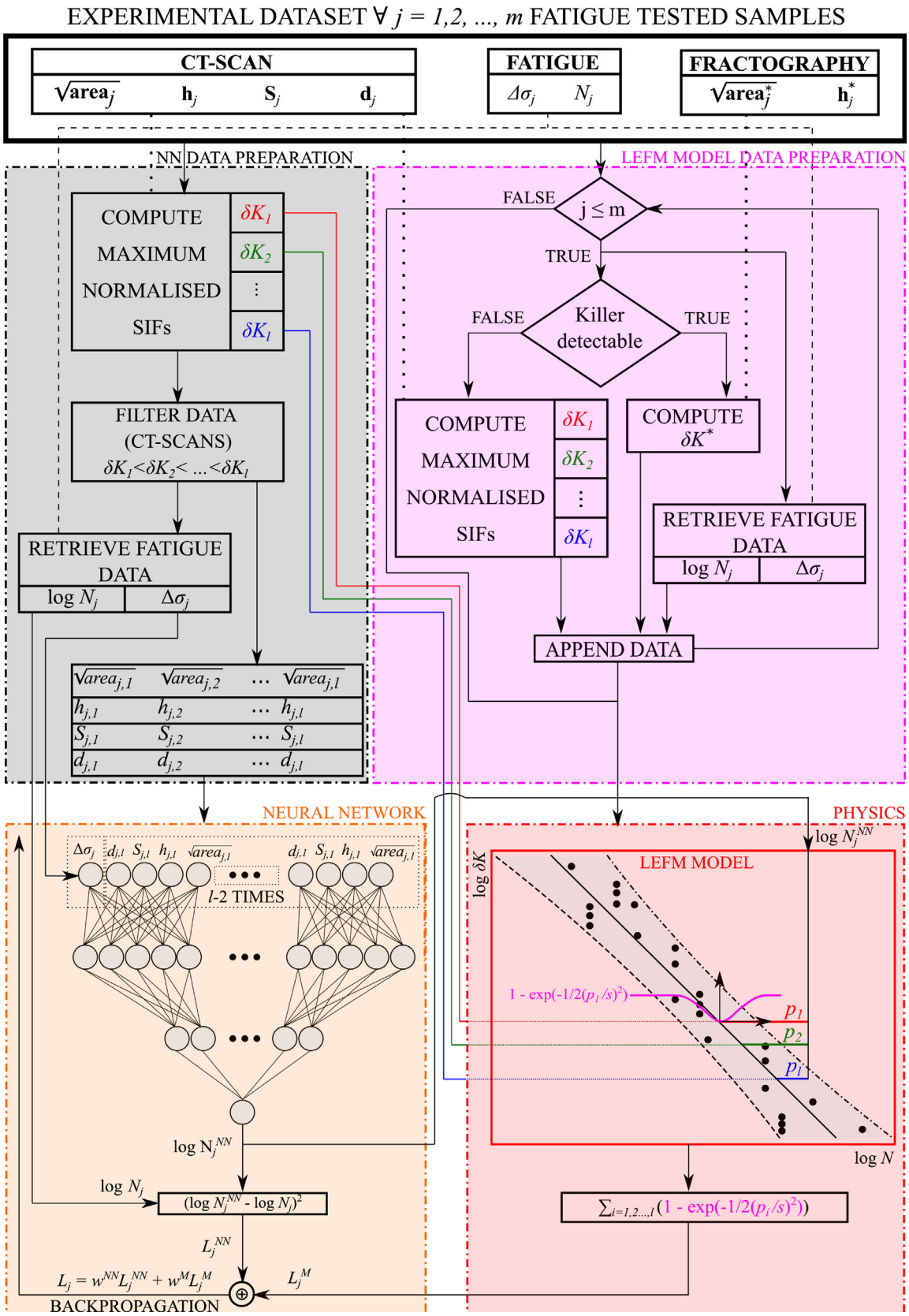


Fig. 3. Developed fatigue finite life PINN algorithm. For graphical reasons, p_i represents $logN_j^{NN} - logN_{j,i}^{*}$ according to Eq. (36). Note that $\oplus$ is the weighted sum between L_j^M and L_j^{NN} through the weights w^M and w^{NN}, respectively.

$$L_j^M = \sum_{i=1}^{l} \left[1 - e^{-\frac{1}{2}\left(\frac{\log N_j^{NN} - \log N_{j,i}^*}{s}\right)^2} \right] \tag{36}$$

For the sake of conciseness, the difference $\log N_j^{NN} - \log N_{j,i}^*$ is labelled as p_i in Fig. 3; where the entire numerical calculation scheme is displayed. Subsequently, the weighted sum of Eq. (30) and Eq. (36) was taken to define the total loss:

$$L_j = w^{NN} L_j^{NN} + w^M L_j^M \tag{37}$$

where $w^{NN} \in [0, 1]$ and $w^M \in [0, 1]$ are weights which are meant to sum to 1, and therefore:

$$w^{NN} = 1 - l \cdot w^M \tag{38}$$

L_j was backpropagated to train the PINN, and the training process had been terminated when L_j was sufficiently small.

Considering Eq. (36), it is worth emphasising that the loss provided from the physics side of the PINN, L_j^M, is originated from the mismatch between the predictions given by the NN, $\log N_j^{NN}$, and the expected value obtained from the regression line of the LEFM, $\log N_{j,i}^*$. From a practical standpoint, the additional contribution provided by L_j^M progressively instruct the PINN to make predictions closer to the regression line, until reaching a trade-off with the experimental input. At the end of the training process, the peculiar structure of L_j^M automatically ensure the predictions to lay inside the prediction band, upon rational choice of the weight w^M, thus complying with the semi-empirical law of the LEFM model, Eq. (21).

Aiming at evaluating the benefits and the accuracy of the PINN as compared with those of an equivalent fully NN one, two indicators were employed, namely the Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2). The RMSE was defined as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{n} \left(\log N_j - \log N_j^{NN}\right)^2}{n}} \tag{39}$$

where N_j is the experimental fatigue life, whereas N_j^{NN} is the associate prediction, and n is the sample size. R^2 was defined as:

$$R^2 = 1 - \frac{\sum_{j=1}^{n} \left(\log N_j - \log N_j^{NN}\right)^2}{\sum_{j=1}^{n} \left(\log N_j - \overline{\log N}\right)^2} \tag{40}$$

where $\overline{\log N}$ is the mean of the observed samples:

$$\overline{\log N} = \frac{1}{n} \sum_{j=1}^{n} \log N_j \tag{41}$$

Fig. 3 shown a summary of the calculation scheme proposed in the present study.

2.3.3. Application to a case-study

To showcase the capabilities of the proposed method, a case-study is presented. In order to pursue a more conservative approach, in the present work, it was decided to pick the first three defects showing the highest SIF range, with the difference that the normalised SIF range (δK) was employed herein. Therefore, l = 3 was assumed. This assumption is justified by the observations reported by Romano et al [17] who systematically identified the two defects showing the highest SIF as the killer defects. It is fundamental to highlight that the selection of l is highly dependent on the distribution of defects giving rise to close SIF range values. For instance, if a certain batch of material shows two predominant defects, having much larger SIF ranges as compared with the remaining ones, then the use of $l > 2$ would not be the most appropriate choice.

As reported in Table 1, a dataset of 12 samples was considered in the present paper. The architecture of the NN consists of a 13-element input layer (according to the dimension of the input vector), a 16-neuron hidden layer, an 8-neuron hidden layer and a single-neuron output layer. Sigmoid activation functions were adopted regarding the neurons of both the first and second hidden layers. Additionally, at each neuron the bias parameter was trained according to the considered data fold. The architecture of the NN was accurately designed to provide predictions complying with the LEFM model, while at the same time keeping its complexity as low as possible and avoiding overfitting the training dataset. It should be noted, however, that no rigorous procedures exist to design the architecture of the NN and thus the necessary number of layers and neurons.

In the present study, some criteria were considered while structuring the NN architecture. The fundamental idea is to gradually increase the complexness of the NN structure. Therefore, preliminary a structure without any hidden layers was tested to check whether the system could be predicted by a simple model. This operation was done by considering R^2, and it was decided to increase the complexity of the NN architecture (by adding more hidden layers) such that the R^2 turned out to be around 0.9 when considering both training and test datasets in the PINN framework; higher values were not considered to avoid overfitting. In the same way, also the number of neurons in the hidden layers were carefully chosen.

The numerical implementation of the PINN was carried out by using *PyTorch* [64]. After preliminary tests, n.2000 epochs was identified as the most suitable value for the present work. During the training process the loss function in Eq. (34), was optimised by means of Adam, a popular and broadly employed Gradient Descent optimiser [65].

It is worth noting that in small data regime, L2-regularisation and adaptive Learning Rate (LR) might be considered to avoid overfitting effects [66 67]. The LR is a specific parameter of the NN which establishes the step size of a Gradient Descent optimiser when exploring the domain of the loss function to be minimised [66]. On one hand, higher LR can accelerate the training process, but on the other, it could trap the optimiser at a local minimum, thus preventing the model from refining the predictions of the unseen data. Furthermore, the training of the NN could benefit from a limited LR, particularly in the small dataset regime; this would make the training process computationally more expensive though. The choice of the LR should also be determined in agreement with the number of training epochs, so a trade-off between all the listed competing factors should be sought.

According to the adopted number of epochs, an adaptive LR was defined by leveraging the best of both sides of the LR spectrum while maintaining an LR globally small to comply with the small dataset regime. Specifically, the initial LR was fixed to 0.001 and held constant until the 1500th epoch. Following, the LR is reduced by 25% every 25 epochs until the 1600th epoch and finally reduced by 75% every 15 epochs. Alongside, L2 regularization with a weight decay as high as 10^{-5} was adopted to further prevent overfitting.

3. Results and discussion

3.1. Applicability of SIF range as a fatigue driving force

Before showing the key results of the present study, it is important to assess the suitability of the proposed fatigue driving force

parameter, i.e. the normalised SIF range δK. In order to do so, the whole set of fatigue life experimental results were considered and fitted using the power law early referred to as the $\delta K - N$ trend (i.e. Eq. (19)), not only for that specific purpose, but also to estimate the $S - N$ and $\Delta K - N$ regression curves, see Fig. 4. In this particular case study, some of the samples were analysed through microscopy to identify the actual defect that triggered the fatigue failure. In those cases, it is unnecessary to consider the first l defects for the regression of the $\delta K - N$ curve and its prediction band. Therefore, in the regression process, a single value of $\delta K - N$ was employed for those samples and expanded l times to give the same weight of the samples in which the killer defect was not identified.

The square root of the estimator for the variance (s) associated with log$N$, concerning the prediction band, referring to Eq. (22), was evaluated to quantify the scatter of each dataset and assess the buoyancy of the regression. As shown in Fig. 4(a), when considering the applied stress range $\Delta\sigma$ as a driving force, the regression provides a good representation of the fatigue behaviour and a prediction band width $s = 0.7$. Instead, this scatter turned out to be significantly higher if the sole ΔK is considered ($s = 1.079$), in contrast with what was reported by Sheridan in his results [47,57], see Fig. 4(b). This contrasting result can be due to the dimension of defects that lay within a region where the role of the non-constant ΔK_{th} is relevant. Nonetheless, if the normalised SIF (δK) is deemed to be a representative driving force, the scatter becomes significantly smaller Fig. 4(c) – although, again, not strikingly evident as showed by Sheridan. Therefore, it can be confidently stated that the normalisation approach implemented herein for the first time can provide a more truthful fatigue finite life representation.

3.2. PINN predictions and K-fold cross validation

As mentioned earlier, a K-fold validation test was performed to check the robustness of the presented approach. This test was successfully carried out for 6 different permutations of the training and test datasets (see Fig. 5(a)-(f)).

As far as the training set is concerned, the blue circles in Fig. 5 are the experimental data, whereas the blue crosses markers are the corresponding predictions. While, the test set is labelled using red circles and crosses, respectively for the experimental and associated predictions. The grey-filled region is the prediction band (Eq. (21)) enclosed between the associated upper (dot-dashed line) and lower limits (dashed line), respectively. The black solid line is the regression curve obtained from the training dataset through OLS, according to Eq. (20). Additionally, each figure reports the square root of the estimator for the variance computed from the portion of data of the training dataset used to build the regression model (Eq. (22)). It is very important to mention again that both the prediction band and the regression line are defined using the known experimental data of the fatigued samples employed as training samples, with δK calculated from fractography if present, otherwise from CT scans as schematically illustrated in Fig. 3; priority must be given to the fractography data since this information pro-

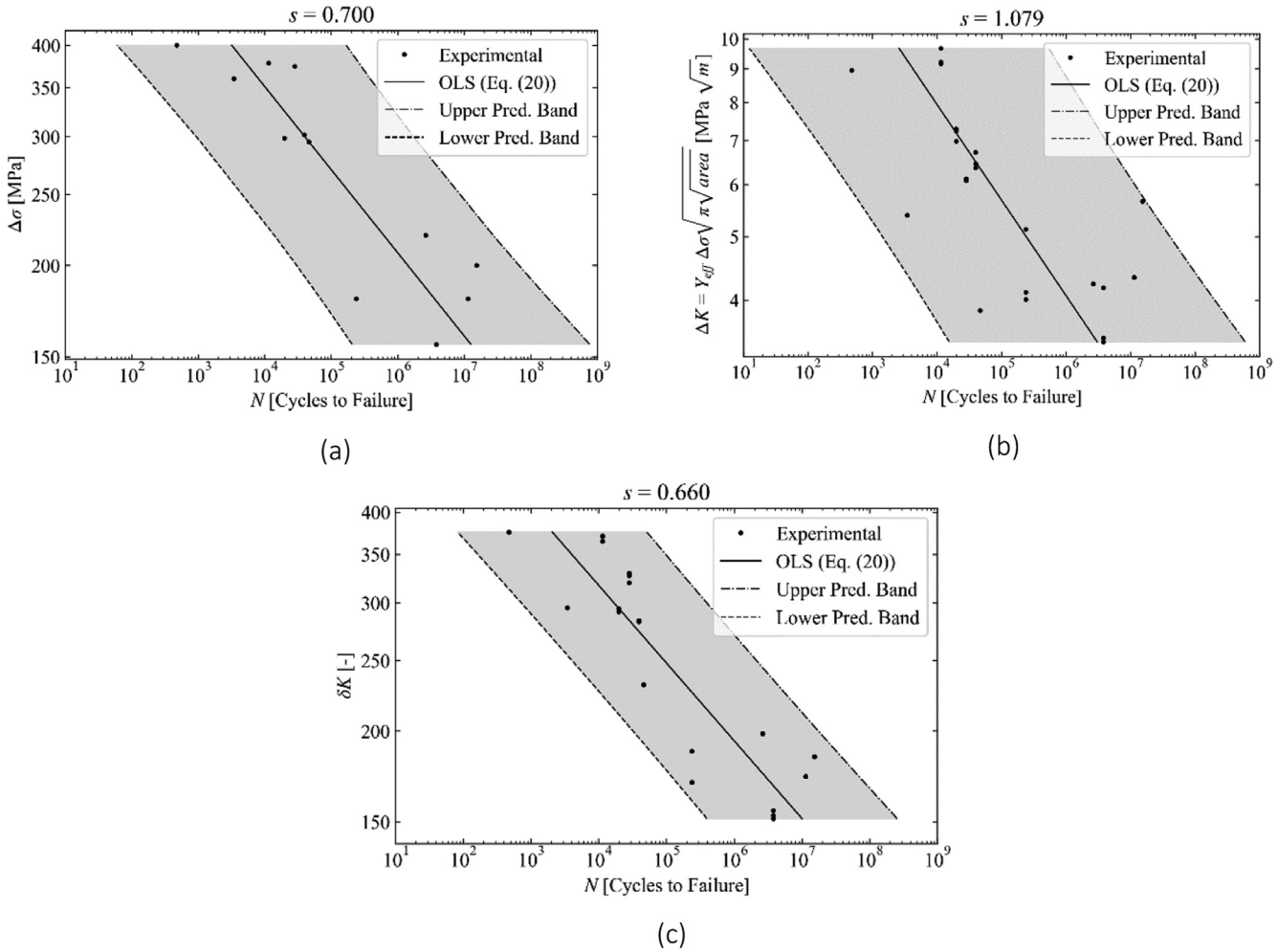


Fig. 4. Regression curves of fatigue life, prediction bands and square root of the estimator for the variance (s) at 95% confidence level. (a) $S - N$; (b) $\Delta K - N$; (c).$\delta K - N$.

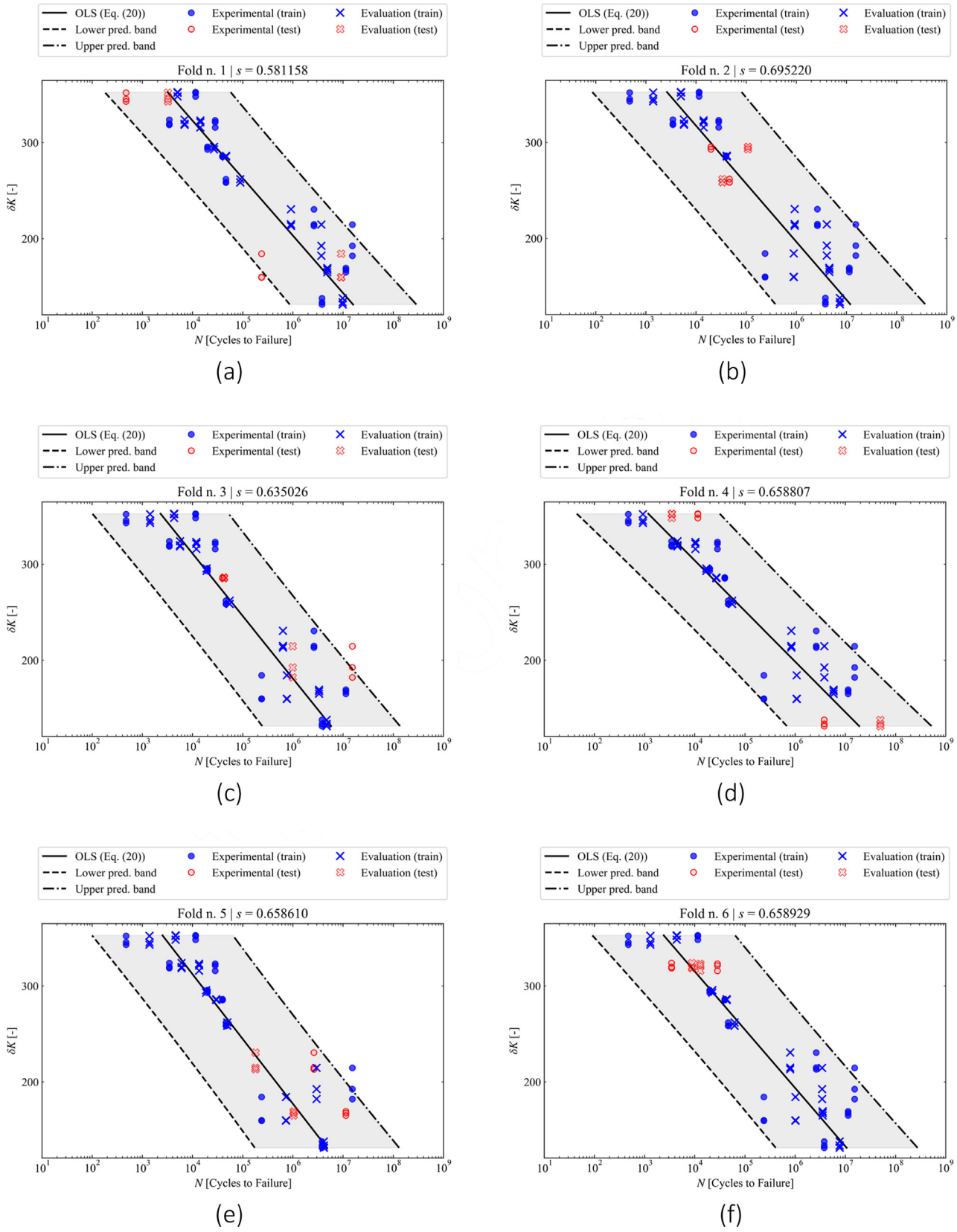


Fig. 5. Predicted fatigue life overlapped onto the prediction band associated with the considered test samples. Different permutations between test and validations samples are shown (K-fold) (a-f).

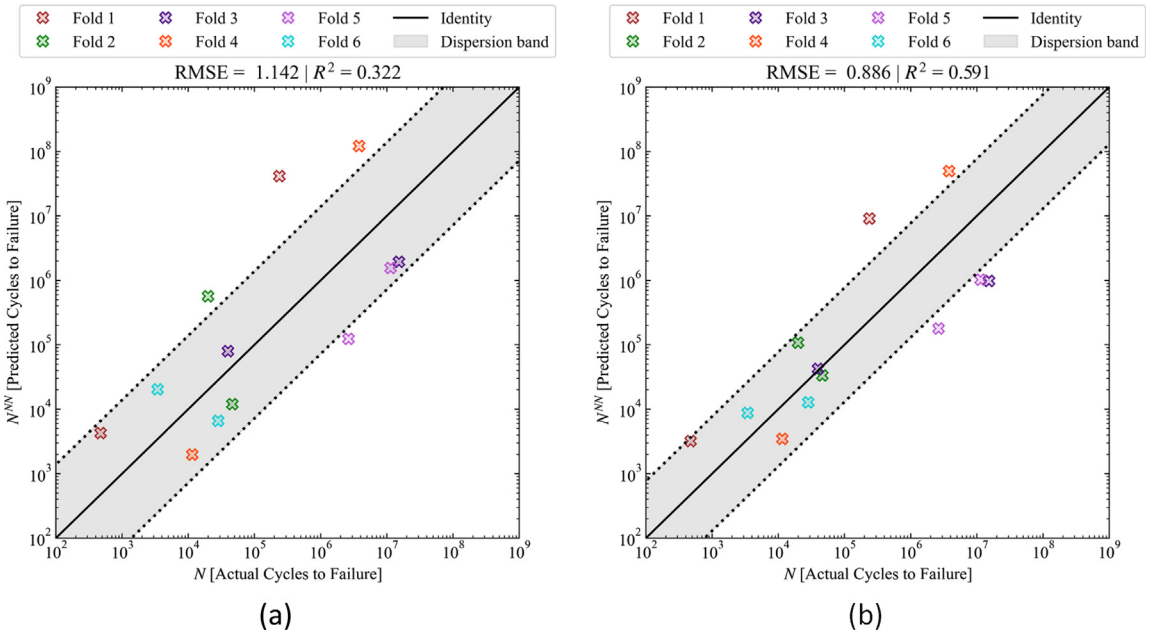


Fig. 6. Accuracy of the PINN and NN frameworks evaluated through RMSE and R^2. Predicted vs. actual number of cycles for (a) NN and (b) PINN. Dispersion band evaluated at 68% confidence level.

vides a precise identification of the killer defect's traits. On the other hand, the PINN is employed to make predictions exclusively using the data obtained from CT.

Fig. 5(a)-(f) illustrates the results of the PINN obtained by setting $w^M = 0.015$ and $w^{NN} = 0.985$, which sums to 1 in agreement with Eq. (38). These weight values turned out to be appropriate since they equate the contribution of L_j^M and L_j^{NN}. As concerns the early stages of the training a small value of w^M could be particularly suitable. In this instance, the predictions given by the PINN could be forecasted far away from the prediction band thus leading excessive values of L_j^M as compared with those of L_j^{NN}.

Although the formulation of the loss function is arbitrary, its peculiar structure defined in Eq. (38) allows the user to promptly tune the importance of the physics with respect to the pure NN prediction and vice versa. Extreme instances may be seen if $w^M = 0$, then $w^{NN} = 1$, meaning that the PINN reduces to a standard NN, since the physical constraint is deactivated. Conversely, if $w^M = 1/l$, then $w^{NN} = 0$, therefore the predictive frameworks rely exclusively on the physical constraints.

Another important strength that is worth highlighting is that in principle, when the PINN is used for prediction purposes, the outcome can be promptly checked with the physics constraints to verify if it complies with the expected range of fatigue life. To this end, it suffices to ascertain whether the predicted point lays inside the prediction band. If so, this point should be accepted. Conversely, it should be rejected.

3.3. Assessing the performance of the PINN

As shown in Fig. 5(a)-(f), the predictions of the PINN framework and their compliance with phenomenological laws are verified. In fact, the markers indicating the predicted life (i.e. evaluation) lay within the prediction band in all the studied cases. Special attention must be paid to those predictions associated with the "test" samples – those not involved while conducting the PINN training process – which again showed a good prediction. This means that the PINN is capable of satisfactorily estimating the fatigue performance based on the morphological traits and distribution of "killer" defects in the material. Minor mismatches in terms of the number of cycles (actual vs. predicted) are certainly attributed to those mechanisms involved that are completely neglected in the present study, residual stress above all. And it is actually for this reason that the employment of semi-empirical law can ensure realistic predictions, rather than seeking the perfect match in terms of the number of cycles which would be impractical if no prior information regarding residual stress (for example) is known.

It is important to observe that the predictions of the training dataset appear to be located at an intermediate position between the corresponding experimental point and the expected value on the regression line. This could be attributed to the PINN achieving an appropriate trade-off between the model loss L_j^M and the numerical loss L_j^{NN} during the training process.

For the sake of comparison, the reader can find the results obtained by deactivating the physical constrains in the supplementary material, so by employing the standard NN training process, i.e. by setting $w_l = 0$ and $w_M = 1$ while retaining the same architecture – neurons, layers, and activation functions. In these results it can be seen that some points actually lay outside the prediction band, and the estimate valued of the test samples was less effective; this will be shown shortly.

Besides the qualitative analysis just discussed, it essential to discuss the advantages offered by the PINN over the classic NN in a quantitative manner as well, although one method does not exclude the other. To this end, the results of Fig. 5 are conveyed into the equivalent, yet concise representation of Fig. 6, which shows the predicted fatigue life vs the experimental fatigue life exclusively for the test data since these are the independent predictions. Specifically, Fig. 6(a) report the result concerning the dataset given by the sole NN, whereas Fig. 6(b) illustrates that given by the PINN counterpart. Additionally, each figure reports the RMSE (Eq. (39)) and R^2 (Eq. (40)), along with its associated dispersion band for each dataset. The Physics Informed side of the PINN provided fundamental and undoubtedly positive support to the NN learning process and avoided overfitting. A remarkable evidence, showing the superior predictive capabilities of the PINN over the NN, can be seen by comparing the values of RMSE and

R^2, and obviously the narrower scatter band at the same confidence level (68%).

It worth remarking that in case of even smaller data regime, additional support might be introduced by increasing L2-regularisation or decreasing the LR, to limit overfitting.

4. Conclusions

In the present work, the development and implementation of a ML-based prediction tool, constrained by an original semi-empirical laws of fatigue finite life, has demonstrated to be highly effective for the accurate prediction of finite fatigue life performance in materials containing flaws. The key results of the research can be summarised as follows:

- The validation executed by exploiting experimental results (AlSi10Mg) showed how this approach can account for defect features that could not be taken into account otherwise: defect sphericity and three-dimensional equivalent diameter. The predictions as compared with a purely NN-based predictive tool are improved as demonstrated by the R^2 index that increased by around 83%.
- The proposed predictive tool can be easily implemented to consider other characteristics of defects not contemplated in the present work that are thought to play a role in fatigue, for example: eccentricity, angularity, solidity etc. [37]. Therefore, the accuracy of the method can be further increased, provided that large datasets are available.
- The fundamental idea of using the prediction band, instead of relying upon a deterministic reference curve, allows for the model to intrinsically account for other sources of uncertainties that cannot be evaluated, at least in this study, residual stress for instance; residual stresses that are certainly present in the analysed samples and may play a significant role.
- The PINN framework has proven to be particularly suited for those problems where the dataset is not sufficiently large for reliable prediction using a pure Machine Learning technique. Indeed, the lack of data is overcome by introducing phenomenological constraints, i.e. the LEFM model, capable of guaranteeing the physical soundness of predicted outcomes.

In conclusion, the PINN framework introduced in this paper has demonstrated an extraordinary capability to make the best out of the two key approaches in fatigue life assessment, namely semi-empirical laws and machine learning methodologies. This work will pave the way for a new class of predictive tools with unprecedented accuracy and great potential for future developments.

CRediT authorship contribution statement

Enrico Salvati: Conceptualization, Methodology, Investigation, Writing – original draft, Supervision, Funding acquisition. **Alessandro Tognan:** Validation, Investigation, Methodology, Data curation, Conceptualization. **Luca Laurenti:** Methodology. **Marco Pelegatti:** Validation, Methodology, Conceptualization. **Francesco De Bona:** Investigation, Methodology.

Data availability

Data will be made available on request.

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.

Acknowledgements

Simone Romano and Stefano Beretta are acknowledged for providing the experimental data.

Data availability

The raw data required to reproduce these findings are available upon request.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.matdes.2022.111089.

References

[1] J. Liu, A.T. Gaynor, S. Chen, Z. Kang, K. Suresh, A. Takezawa, L. Li, J. Kato, J. Tang, C.C.L. Wang, L. Cheng, X. Liang, A.C. To, Current and future trends in topology optimization for additive manufacturing, Struct. Multidiscip. Optim. 57 (6) (2018) 2457–2483.
[2] N. Sanaei, A. Fatemi, Defects in additive manufactured metals and their effect on fatigue performance: A state-of-the-art review, Prog. Mater Sci. 117 (2021) 100724.
[3] J.P. Oliveira, T.G. Santos, R.M. Miranda, Revisiting fundamental welding concepts to improve additive manufacturing: From theory to practice, Prog. Mater Sci. 107 (2020) 100590.
[4] E. Salvati, A.J.G. Lunt, C.P. Heason, G.J. Baxter, A.M. Korsunsky, An analysis of fatigue failure mechanisms in an additively manufactured and shot peened IN 718 nickel superalloy, Mater. Des. 191 (2020) 108605.
[5] E. Salvati, A.M. Korsunsky, An analysis of macro- and micro-scale residual stresses of Type I, II and III using FIB-DIC micro-ring-core milling and crystal plasticity FE modelling, Int. J. Plast. 98 (2017) 123–138.
[6] J. Everaerts, E. Salvati, F. Uzun, L. Romano Brandt, H. Zhang, A.M. Korsunsky, Separating macro- (Type I) and micro- (Type II+III) residual stresses by ring-core FIB-DIC milling and eigenstrain modelling of a plastically bent titanium alloy bar, Acta Mater. 156 (2018) 43–51.
[7] X. Song, S. Feih, W. Zhai, C.-N. Sun, F. Li, R. Maiti, J. Wei, Y. Yang, V. Oancea, L. Romano Brandt, A.M. Korsunsky, Advances in additive manufacturing process simulation: residual stresses and distortion predictions in complex metallic components, Mater. Des. 193 (2020) 108779.
[8] X. Cui, S. Zhang, C. Wang, C.H. Zhang, J. Chen, J.B. Zhang, Effects of stress-relief heat treatment on the microstructure and fatigue property of a laser additive manufactured 12CrNi2 low alloy steel, Mater. Sci. Eng., A 791 (2020) 139738.
[9] E. Salvati, A.J.G. Lunt, S. Ying, T. Sui, H.J. Zhang, C. Heason, G. Baxter, A.M. Korsunsky, Eigenstrain reconstruction of residual strains in an additively manufactured and shot peened nickel superalloy compressor blade, Comput. Methods Appl. Mech. Eng. 320 (2017) 335–351.
[10] E. Maleki, S. Bagherifard, M. Bandini, M. Guagliano, Surface post-treatments for metal additive manufacturing: Progress, challenges, and opportunities, Addit. Manuf. 37 (2021) 101619.
[11] E. Maleki, S. Bagherifard, S.M.J. Razavi, M. Bandini, A. du Plessis, F. Berto, M. Guagliano, On the efficiency of machine learning for fatigue assessment of post-processed additively manufactured AlSi10Mg, Int. J. Fatigue 160 (2022) 106841.
[12] Y. Chen, S.J. Clark, C.L.A. Leung, L. Sinclair, S. Marussi, M.P. Olbinado, E. Boller, A. Rack, I. Todd, P.D. Lee, In-situ Synchrotron imaging of keyhole mode multi-layer laser powder bed fusion additive manufacturing, Appl. Mater. Today 20 (2020) 100650.
[13] C.L.A. Leung, S. Marussi, R.C. Atwood, M. Towrie, P.J. Withers, P.D. Lee, In situ X-ray imaging of defect and molten pool dynamics in laser additive manufacturing, Nat. Commun. 9 (1) (2018) 1355.
[14] M. Laleh, A.E. Hughes, S. Yang, J. Wang, J. Li, A.M. Glenn, W. Xu, M.Y. Tan, A critical insight into lack-of-fusion pore structures in additively manufactured stainless steel, Addit. Manuf. 38 (2021) 101762.
[15] H. Masuo, Y. Tanaka, S. Morokoshi, H. Yagura, T. Uchida, Y. Yamamoto, Y. Murakami, Influence of defects, surface roughness and HIP on the fatigue strength of Ti-6Al-4V manufactured by additive manufacturing, Int. J. Fatigue 117 (2018) 163–179.
[16] S. Beretta, S. Romano, A comparison of fatigue strength sensitivity to defects for materials manufactured by AM or traditional processes, Int. J. Fatigue 94 (2017) 178–191.

[17] S. Romano, A. Brückner-Foit, A. Brandão, J. Gumpinger, T. Ghidini, S. Beretta, Fatigue properties of AlSi10Mg obtained by additive manufacturing: Defect-based modelling and prediction of fatigue strength, Eng. Fract. Mech. 187 (2018) 165–189.
[18] Y.N. Hu, S.C. Wu, P.J. Withers, J. Zhang, H.Y.X. Bao, Y.N. Fu, G.Z. Kang, The effect of manufacturing defects on the fatigue life of selective laser melted Ti-6Al-4V structures, Mater. Des. 192 (2020) 108708.
[19] G. Meneghetti, D. Rigon, C. Gennari, An analysis of defects influence on axial fatigue strength of maraging steel specimens produced by additive manufacturing, Int. J. Fatigue 118 (2019) 54–64.
[20] S. Romano, L. Patriarca, S. Foletti, S. Beretta, LCF behaviour and a comprehensive life prediction model for AlSi10Mg obtained by SLM, Int. J. Fatigue 117 (2018) 47–62.
[21] Y. Murakami, Metal fatigue: Effects of small defects and nonmetallic inclusions, Elsevier, 2019.
[22] S. Tammas-Williams, P.J. Withers, I. Todd, P.B. Prangnell, The influence of porosity on fatigue crack initiation in additively manufactured titanium components, Sci. Rep. 7 (1) (2017) 7308.
[23] S. Beretta, More than 25 years of extreme value statistics for defects: Fundamentals, historical developments, recent applications, Int. J. Fatigue 151 (2021) 106407.
[24] S. Romano, S. Miccoli, S. Beretta, A new FE post-processor for probabilistic fatigue assessment in the presence of defects and its application to AM parts, Int. J. Fatigue 125 (2019) 324–341.
[25] M.H. El Haddad, K.N. Smith, T.H. Topper, Fatigue Crack Propagation of Short Cracks, J. Eng. Mater. Technol. 101 (1) (1979) 42–46.
[26] A. Wormsen, B. Sjödin, G. Härkegård, A. Fjeldstad, Non-local stress approach for fatigue assessment based on weakest-link theory and statistics of extremes, Fatigue Fract. Eng. Mater. Struct. 30 (12) (2007) 1214–1227.
[27] A. Wormsen, A Fatigue Assessment Methodology for Notched Components Containing Defects, 2007.
[28] M.P. Enright, R.C. McClung, K.S. Chan, J. McFarland, J.P. Moody, J.C. Sobotka, Micromechanics-based fracture risk assessment using integrated probabilistic damage tolerance analysis and manufacturing process models, ASME Turbo Expo 2016: Turbomachinery Technical Conference and Exposition (2016).
[29] S. Romano, A. Abel, J. Gumpinger, A.D. Brandão, S. Beretta, Quality control of AlSi10Mg produced by SLM: metallography versus CT scans for critical defect size assessment, Addit. Manuf. 28 (2019) 394–405.
[30] P. Livieri, F. Segala, Asymptotic behaviour of the Oore-Burns integral for cracks with a corner and correction formulae for embedded convex defects, Eng. Fract. Mech. 252 (2021) 107663.
[31] V. Lazarus, Brittle fracture and fatigue propagation paths of 3D plane cracks under uniform remote tensile loading, Int. J. Fract. 122 (1) (2003) 23–46.
[32] P. Livieri, F. Segala, Three-dimensional fatigue crack propagation by means of first order SIF approximation, Procedia Struct. Integrity 39 (2022) 194–203.
[33] X.-C. Zhang, J.-G. Gong, F.-Z. Xuan, A deep learning based life prediction method for components under creep, fatigue and creep-fatigue conditions, Int. J. Fatigue 148 (2021) 106236.
[34] A. Karolczuk, M. Słoński, Application of the Gaussian process for fatigue life prediction under multiaxial loading, Mech. Syst. Sig. Process. 167 (2022) 108599.
[35] J. Li, Z. Yang, G. Qian, F. Berto, Machine learning based very-high-cycle fatigue life prediction of Ti-6Al-4V alloy fabricated by selective laser melting, Int. J. Fatigue 158 (2022) 106764.
[36] H. Bao, S. Wu, Z. Wu, G. Kang, X. Peng, P.J. Withers, A machine-learning fatigue life prediction approach of additively manufactured metals, Eng. Fract. Mech. 242 (2021) 107508.
[37] A. Li, S. Baig, J. Liu, S. Shao, N. Shamsaei, Defect criticality analysis on fatigue life of L-PBF 17–4 PH stainless steel via machine learning, Int. J. Fatigue 163 (2022) 107018.
[38] L. He, Z. Wang, Y. Ogawa, H. Akebono, A. Sugeta, Y. Hayashi, Machine-learning-based investigation into the effect of defect/inclusion on fatigue behavior in steels, Int. J. Fatigue 155 (2022) 106597.
[39] X. Peng, S. Wu, W. Qian, J. Bao, Y. Hu, Z. Zhan, G. Guo, P.J. Withers, The potency of defects on fatigue of additively manufactured metals, Int. J. Mech. Sci. 221 (2022) 107185.
[40] S. Moon, R. Ma, R. Attardo, C. Tomonto, M. Nordin, P. Wheelock, M. Glavicic, M. Layman, R. Billo, T. Luo, Impact of surface and pore characteristics on fatigue life of laser powder bed fusion Ti–6Al–4V alloy described by neural network models, Sci. Rep. 11 (1) (2021) 20424.
[41] G.E. Karniadakis, I.G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, L. Yang, Physics-informed machine learning, Nat. Rev. Phys. 3 (6) (2021) 422–440.
[42] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707.
[43] R. Liu, S. Liu, X. Zhang, A physics-informed machine learning model for porosity analysis in laser powder bed fusion additive manufacturing, Int. J. Adv. Manufact. Technol. 113 (7) (2021) 1943–1958.
[44] J. Chen, Y. Liu, Fatigue property prediction of additively manufactured Ti-6Al-4V using probabilistic physics-guided learning, Addit. Manuf. 39 (2021) 101876.
[45] C.-T. Chen, X. Gu Grace, Learning hidden elasticity with deep neural networks, Proceedings of the National Academy of Sciences 118(31) (2021) e2102721118.
[46] H. Masuo, Y. Tanaka, S. Morokoshi, H. Yagura, T. Uchida, Y. Yamamoto, Y. Murakami, Effects of defects, surface roughness and HIP on fatigue strength of Ti-6Al-4V manufactured by additive manufacturing, Procedia Struct. Integrity 7 (2017) 19–26.
[47] L. Sheridan, J.E. Gockel, O.E. Scott-Emuakpor, Stress-defect-life interactions of fatigued additively manufactured alloy 718, Int. J. Fatigue 143 (2021) 106033.
[48] Z. Wu, S. Wu, J. Bao, W. Qian, S. Karabal, W. Sun, P.J. Withers, The effect of defect population on the anisotropic fatigue resistance of AlSi10Mg alloy fabricated by laser powder bed fusion, Int. J. Fatigue 151 (2021) 106317.
[49] K. Solberg, S. Guan, S.M.J. Razavi, T. Welo, K.C. Chan, F. Berto, Fatigue of additively manufactured 316L stainless steel: the influence of porosity and surface roughness, Fatigue Fract. Eng. Mater. Struct. 42 (9) (2019) 2043–2052.
[50] Y. Murakami, T. Takagi, K. Wada, H. Matsunaga, Essential structure of S-N curve: Prediction of fatigue life and fatigue limit of defective materials and nature of scatter, Int. J. Fatigue 146 (2021) 106138.
[51] Y. Yamashita, Y. Murakami, Small crack growth model from low to very high cycle fatigue regime for internal fatigue failure of high strength steel, Int. J. Fatigue 93 (2016) 406–414.
[52] S. Suresh, R.O. Ritchie, Propagation of short fatigue cracks, Int. Met. Rev. 29 (1) (1984) 445–473.
[53] M.D. Chapetti, Fatigue propagation threshold of short cracks under constant amplitude loading, Int. J. Fatigue 25 (12) (2003) 1319–1326.
[54] B. Künkler, O. Düber, P. Köster, U. Krupp, C.P. Fritzen, H.J. Christ, Modelling of short crack propagation – transition from stage I to stage II, Eng. Fract. Mech. 75 (3) (2008) 715–725.
[55] E. Salvati, H. Zhang, K.S. Fong, X. Song, A.M. Korsunsky, Separating plasticity-induced closure and residual stress contributions to fatigue crack retardation following an overload, J. Mech. Phys. Solids 98 (2017) 222–235.
[56] Y. Murakami, M. Endo, Effects of Hardness and Crack Geometries on ΔKth of Small Cracks Emanating from Small Defects, Mechanical Engineering Publications, The Behaviour of Short Fatigue Cracks, 1986, pp. 275–293.
[57] L. Sheridan, A modified El-Haddad model for versatile defect tolerant design, Int. J. Fatigue 145 (2021) 106062.
[58] E. Astm, 739–91. Standard practice for statistical analysis of linear or linearized stress-life (SN) and strain-life (ε-N) fatigue data, ASTM Int. (2015).
[59] S. De Gryze, I. Langhans, M. Vandebroek, Using the correct intervals for prediction: a tutorial on tolerance intervals for ordinary least-squares regression, Chemomet. Intell. Laborat. Syst. 87 (2) (2007) 147–154.
[60] P. Paris, F. Erdogan, A critical analysis of crack propagation laws, J. Basic Eng. 85 (4) (1963) 528–533.
[61] E. Salvati, T. Sui, H. Zhang, A.J.G. Lunt, K.S. Fong, X. Song, A.M. Korsunsky, Elucidating the mechanism of fatigue crack acceleration following the occurrence of an underload, Adv. Eng. Mater. 18 (12) (2016) 2076–2087.
[62] F. Kun, H.A. Carmona, J.S. Andrade, H.J. Herrmann, Universality behind Basquin's law of fatigue, Phys. Rev. Lett. 100 (9) (2008) 094301.
[63] M.E. Kassner, M.T. Pérez-Prado, Five-power-law creep in single phase metals and alloys, Prog. Mater Sci. 45 (1) (2000) 1–102.
[64] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, Pytorch: An imperative style, high-performance deep learning library, Adv. Neural Inform. Process. Syst. 32 (2019).
[65] Y. Wu, L. Liu, J. Bae, K.-H. Chow, A. Iyengar, C. Pu, W. Wei, L. Yu, Q. Zhang, Demystifying learning rate policies for high accuracy training of deep neural networks,, IEEE International conference on big data (Big Data), IEEE 2019 (2019) 1971–1980.
[66] L.N. Smith, A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay, arXiv preprint arXiv:1803.09820 (2018).
[67] T. Van Laarhoven, L2 regularization versus batch and weight normalization, arXiv preprint arXiv:1706.05350 (2017).